

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2002, 10:21:35 ; Search time 1830.12 Seconds
(without alignments)
10565.493 Million cell updates/sec

Title: US-09-975-308-8
Perfect score: 924

Sequence: 1 atgaacacagcgctgttaac.....ttgcatttcgaacactag 924

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba:*
2: gb_bt:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pt:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgc_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query Score Match Length DB ID Description

C	1	922.4	99.8	180657	2	AC091612	AC091612 Homo sapi
	2	920.8	99.7	138591	2	AC026038	AC026038 Homo sapi
	3	920.8	99.7	190889	2	AL357039	AL357039 Homo sapi
	4	917.8	99.3	921	6	AX241862	AX241862 Sequence
	5	778.8	84.3	798	6	AX350633	AX350633 Sequence
C	6	687.8	74.4	221466	2	AC099601	AC099601 Mus muscu
	7	666.6	72.1	160483	2	AC094718	AC094718 Rattus no
C	8	665.6	72.0	926	10	AY073900	AY073900 Rattus no
	9	661.4	71.6	663	6	AX350675	AX350675 Sequence
	10	481.4	52.1	59688	2	AC101272	AC101272 Mus muscu
	11	370.2	40.1	916	10	AY074180	AY074180 Mus muscu
	12	369.2	40.0	915	10	AY073135	AY073135 Mus muscu
C	13	339	36.7	112730	2	AC094870	AC094870 Rattus no
	14	339	36.7	197438	2	AC094700	AC094700 Rattus no
C	15	331.2	35.8	204143	2	AC096461	AC096461 Rattus no
	16	326.2	35.3	930	10	AY073132	AY073132 Mus muscu
	17	323.2	35.0	930	10	AY073176	AY073176 Mus muscu
	18	319	34.5	930	10	AY073062	AY073062 Mus muscu
	19	319	34.5	116603	2	AC094493	AC094493 Rattus no
	20	317.2	34.3	936	10	AY073133	AY073133 Mus muscu
	21	310.8	33.6	939	10	AY073126	AY073126 Mus muscu
	22	309.2	33.5	936	10	AY073125	AY073125 Mus muscu
	23	307.6	33.3	939	10	AY073128	AY073128 Mus muscu
	24	307.2	33.2	1241	6	AX207662	AX207662 Sequence
C	25	307.2	33.2	160654	2	AC011879	AC011879 Homo sapi
	26	307.2	33.2	168473	9	AL512324	AL512324 Human DNA
	27	306	33.1	960	10	AY073127	AY073127 Mus muscu
	28	305.4	33.1	924	6	AX241531	AX241531 Sequence
	29	304	32.9	957	10	AY073154	AY073154 Mus muscu
	30	302.2	32.7	197438	2	AC094700	AC094700 Rattus no
	31	295.4	32.0	936	10	AY073124	AY073124 Mus muscu
	32	292.8	31.7	939	10	AY073134	AY073134 Mus muscu
	33	289.6	31.3	112730	2	AC094870	AC094870 Rattus no
	34	281.6	30.5	924	10	AY073578	AY073578 Mus muscu
	35	280	30.3	933	10	AY073073	AY073073 Mus muscu
	36	278.4	30.1	924	10	AY073577	AY073577 Mus muscu
	37	274.2	29.7	763	10	AY073895	AY073895 Mus muscu
	38	274	29.7	928	10	AY074052	AY074052 Mus muscu
	39	274	29.7	95794	2	AC108568	AC108568 Rattus no
	40	269.2	29.1	948	10	AY073382	AY073382 Mus muscu
C	41	268.2	29.1	203356	2	AL669952	AL669952 Mus muscu
C	42	268.4	29.0	116603	2	AC094493	AC094493 Rattus no
	43	267.4	28.9	972	6	AX241860	AX241860 Sequence
	44	267.4	28.9	1050	6	AX320287	AX320287 Sequence
C	45	267.4	28.9	156604	2	AC069371	AC069371 Homo sapi

ALIGNMENTS

RESULT 1
AC091612/c
LOCUS AC091612
DEFINITION Homo sapiens chromosome 1 clone RP11-656022, WORKING DRAFT
ACCESSION AC091612 AL390860
VERSION AC091612.4 GI:18497169
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 180657)
AUTHORS Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 180657)
AUTHORS Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J., Ivey,R.G. and Haugen,E.D.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
COMMENT On Feb 5, 2002 this sequence version replaced gi:15487406.

Chemistry: Dye-terminator Big Dye; 0% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 129539 bases at least Q40
 Consensus quality: 131919 bases at least Q30
 Consensus quality: 133309 bases at least Q20
 Insert size: 140000; agarose-fp
 Insert size: 136491; sum-of-contigs
 Quality coverage: 3.88 in Q20 bases; agarose-fp
 Quality coverage: 4.04 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 22 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1      5221: contig of 5221 bp in length
*      5222      5321: gap of unknown length
*      5322      9571: contig of 4250 bp in length
*      9572      13673: gap of unknown length
*      13674      13773: gap of unknown length
*      13774      18721: contig of 4948 bp in length
*      18722      18821: gap of unknown length
*      18822      24023: contig of 5202 bp in length
*      24024      24123: gap of unknown length
*      24124      28988: contig of 4865 bp in length
*      28989      29088: gap of unknown length
*      29089      34171: contig of 5083 bp in length
*      34172      34271: gap of unknown length
*      34272      40989: contig of 6718 bp in length
*      40990      41089: gap of unknown length
*      41090      48353: contig of 7266 bp in length
*      48356      48455: gap of unknown length
*      48456      59366: contig of 10911 bp in length
*      59367      59466: gap of unknown length
*      59467      61313: contig of 1847 bp in length
*      61314      61413: gap of unknown length
*      61414      71241: contig of 9828 bp in length
*      71242      80867: contig of 9526 bp in length
*      80868      80967: gap of unknown length
*      80968      99688: contig of 18721 bp in length
*      99689      99789: gap of unknown length
*      99790      122864: contig of 23076 bp in length
*      122865      122964: gap of unknown length
*      122965      124804: contig of 1840 bp in length
*      124805      124904: gap of unknown length
*      124905      126620: contig of 1716 bp in length
*      126621      126720: gap of unknown length
*      126721      128799: contig of 2079 bp in length
*      128800      128899: gap of unknown length
*      128900      130561: contig of 1662 bp in length
*      130562      130661: gap of unknown length
*      130662      132547: contig of 1886 bp in length
*      132548      132647: gap of unknown length
*      132648      135428: contig of 2781 bp in length
*      135429      135528: gap of unknown length
*      135529      138591: contig of 3063 bp in length.

```

FEATURES

source

1..138591

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="1"

/clone="RP11-109C14"

/note="assembly_name:Contig10"

clone_end:=77

vector_side:left"

misc_feature

/note="assembly_name:Contig11"

```

misc_feature      9672..13673
                  /note="assembly_name:Contig12"
misc_feature      13774..18721
                  /note="assembly_name:Contig13"
misc_feature      18822..24023
                  /note="assembly_name:Contig14"
misc_feature      24124..28988
                  /note="assembly_name:Contig15"
misc_feature      29089..34171
                  /note="assembly_name:Contig16"
misc_feature      34272..40989
                  /note="assembly_name:Contig17"
misc_feature      41090..48355
                  /note="assembly_name:Contig18"
misc_feature      48456..59366
                  /note="assembly_name:Contig19"
misc_feature      59467..61313
                  /note="assembly_name:Contig2"
misc_feature      61414..71241
                  /note="assembly_name:Contig20"
misc_feature      71342..80867
                  /note="assembly_name:Contig21"
misc_feature      80968..99688
                  /note="assembly_name:Contig22"
misc_feature      99789..122864
                  /note="assembly_name:Contig23"
misc_feature      clone_end:=76
                  vector_side:left"
misc_feature      122965..124804
                  /note="assembly_name:Contig3"
misc_feature      124905..126620
                  /note="assembly_name:Contig4"
misc_feature      126721..128799
                  /note="assembly_name:Contig5"
misc_feature      128900..130561
                  /note="assembly_name:Contig6"
misc_feature      130662..132547
                  /note="assembly_name:Contig6"
misc_feature      132648..135428
                  /note="assembly_name:Contig7"
misc_feature      135529..138591
                  /note="assembly_name:Contig8"
misc_feature      /note="assembly_name:Contig9"
BASE COUNT      44211 a 25199 c 25318 g 41754 t 2109 others
ORIGIN
Query Match      99.7% Score 920.8; DB 2; Length 138591;
Best Local Similarity 99.8%; Pred. No. 3.4e-251;
Matches 922; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      1 atgaatcacaggtgtgaactgaattcaattatcttgggagcttaccacaaagcttgaactc 60
DB      86786 ATGAATCACAGGCTGTGAATCTAGTTCAATTATCTGGGCTTACCAAAAAGCTGAATCTC 86845
QY      61 cagggaattatctctctctctctctctctctctctctctctctctctctctctctctctct 120
DB      86846 CAGGGAATTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 86905
QY      121 cccatcatcatctgccaataatcctatagcaaacctctgcatagcgccatgatatgttctcct 180
DB      86906 CTCATCATCATCTGCCAATAATCTATTAACAACCTGTGCATACCCCATATGATGTTTCCCT 86965
QY      181 ctgacactgagctggtgtgagacatctctgcacaaagacatataccgaagatgcttggg 240
DB      86966 CTGACACTGAGCTGTGTGTGACATCTCTGCACAAAGACATATACCGAATGCTTGGGG 87025
QY      241 accatgtctaaacatcagaaataacatctcatatctgagagctgcatgtctccagctctctctg 300
DB      87026 ACCATGCTAAACATCAGAAATACATTCATATGACAGGCTGATGCCAGCTCTTCTTG 87085
QY      301 ttcaatgtctctctgagagctgagatgattctctccacacacatgctctatgacgcctcat 360
DB      87086 TTCAATGTCTCTCTGAGAGCTGAGATGATTCTTTCACACCATGCTATGACCGCTAT 87145

```



```
misc_feature      fragment_chain:2"
                    57909..68643
                    /note="assembly_fragment:01434
                    fragment_chain:2"
misc_feature      68744..75536
                    /note="assembly_fragment:01941
                    fragment_chain:2"
misc_feature      75637..84190
                    /note="assembly_fragment:00841
                    fragment_chain:2"
misc_feature      84291..89039
                    /note="assembly_fragment:00517
                    fragment_chain:3"
misc_feature      89140..94516
                    /note="assembly_fragment:00194
                    fragment_chain:3"
misc_feature      94617..105920
                    /note="assembly_fragment:00685
                    fragment_chain:4"
misc_feature      106021..133564
                    /note="assembly_fragment:00244
                    fragment_chain:4"
misc_feature      133665..137940
                    /note="assembly_fragment:00993
                    fragment_chain:5"
misc_feature      138041..141386
                    /note="assembly_fragment:00049
                    fragment_chain:5"
misc_feature      141487..143513
                    /note="assembly_fragment:01137"
                    fragment_chain:5"
misc_feature      143614..159175
                    /note="assembly_fragment:01189"
                    fragment_chain:5"
misc_feature      159276..174409
                    /note="assembly_fragment:01199"
                    fragment_chain:5"
misc_feature      174510..190889
                    /note="assembly_fragment:01519"
                    fragment_chain:5"
BASE COUNT      60637 a 34972 c 35649 g 57919 t 1712 others
ORIGIN
Query Match      99.7%; Score 920.8; DB 2; Length 190889;
Best Local Similarity 99.8%; Pred. No. 3.5e-251;
Matches 922; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Db 116598 TTGCTACAGTGTGATGCTATGCTACGATCACCATTCTGGGTGACACAGCTTTATTC
|||||
Qy 481 atgaagtgacttctctgtggaacacacattgacacttctctgtgagatacccca 540
|||||
Db 116658 ATGAGGTGACTTCTGTGTGGGCCAACAACACATGACCACTTCTGTGAGATACCCCA 116717
|||||
Qy 541 ttgctgcttctctgtgacccctgtaagaatcaatgagtgatgtatgtctgat 600
|||||
Db 116718 TTGCTGGCTTTGCTGTGAGCCCTGTAGAAATCAAGAGTGTATGTTGCTGAT 116777
|||||
Qy 601 ataaccttgccataagggaacttattcttaactgatatccatggtttacattgt 660
|||||
Db 116778 ATTACCTTGCCATAGGAGACTTTATTCTTACCTGATCTCTATGTTTATCATGTT 116837
|||||
Qy 661 gctatctccglatccgcacagtagaaggcaaggagccctctcaacatgctcatc 720
|||||
Db 116838 GCTATCTCCGATATCCGACAGTAGAAGGCAAGAGAGAGCCCTTCCAAATGCTCATCT 116897
|||||
Qy 721 catctacagtggtgaccttactatctctctgtatcttaacatctccgctgct 780
|||||
Db 116898 CATCTCAGAGTGTGACCCCTTACTATTCTCTGTATGTACACCTATATCCGCTGCT 116957
|||||
Qy 781 tcaagctatcatcttgaagagacaagtgtagctgcactctatactctgtgactccc 840
|||||
Db 116958 TCCAGCTATACATTTCAGAAAGACAGAGGTGTAGCTGCACCTATATCTTGTGACTCC 117017
|||||
Qy 841 acattaaacccagatggtgtacagcttccagaaataggagatgcaggcaataggaa 900
|||||
Db 117018 ACATTAAACCCAGTGTGTACAGCTTCCAGAAATAGGAGATGACAGGCAAGATTAGAA 117077
|||||
Qy 901 gtgtttgcatcttcgaacactag 924
Db 117078 GTGTTGCAATTTCTGAACACTAG 117101
|||||

RESULT 4
AX241862      921 bp      DNA      linear      PAT 26-SEP-2001
LOCUS      AX241862      Sequence 610 from Patent WO0127158.
DEFINITION      AX241862
ACCESSION      AX241862
VERSION      AX241862.1 GI:15798737
KEYWORDS
SOURCE
ORGANISM      synthetic construct.
artificial sequence.
REFERENCE      1 (bases 1 to 921)
AUTHORS      Bellenson,J., Smith,D., Lancel,D., Glusman,G., Fuchs,T. and
Yanai,I.
TITLE      Olfactory receptor sequences
JOURNAL      Patent: WO 0127158-A 610 19-APR-2001.
YEDDA RESEARCH AND DEVELOPMENT COMPANY, LTD. (IL)
FEATURES
source
1..921
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="(H38g459 nucleotide)."
BASE COUNT      218 a 235 c 183 g 285 t
ORIGIN
Query Match      99.3%; Score 917.8; DB 6; Length 921;
Best Local Similarity 99.8%; Pred. No. 1.5e-250;
Matches 919; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

QY 121 ctcatatcatcccaaaatctatagcaaaccttgcatacagcccatgtatgttctt 180
 Db 121 CTTATCATATATTCGCAAAATCTATACACACCTTGCATACGCCCATGTATGTTCCCT 180
 QY 181 ctgacactgctgtgttgagacatctgcacaaagaacataacagaagatgctggg 240
 Db 181 CTGACACTGGCTGTGTGAGCATCTGACACAAACAGCATATACCGAAGATGCTGGGG 240
 QY 241 accatgctaacatcaaaaataacattctatctgacaggtctgacgtctctctg 300
 Db 241 ACCATGCTAACATCAAAAATACCATTTCTATGACAGGCTGCATGCTCCAGCTCTTCTG 300
 QY 301 ttacatgctctctggagctgagatggtctcttcaacacattggcctatagcgtat 360
 Db 301 TTCATAGGCTCTGGAGCTGAGATGCTCTCTTCACCCACATGGGCTATGACCCCTAT 360
 QY 361 gtggccattgttccctcttcattacagtaactatatacgaaccacatctgtgtgagcc 420
 Db 361 GTGGCCATTGTTCCTCCCTTCATTTACAGTACTGTTATGAAACACCATATGTGTGAGCC 420
 QY 421 ttgtctgacatgctcagtgctattgtcagtaaccaatcctcgggtgcacacagctcttacc 480
 Db 421 TTGCTGACATGCTCATAGCTATTGACATCACCATTCTCGGGTGACACAGCTCTTATC 480
 QY 481 atgaggttgaactctctgtggccaacacacattgacacattctctctgtgagatacccca 540
 Db 481 ATGAGGTTGACTTCTGTGGGGCAAAACACATGACACACTTCTCTGTGAGATACCCCA 540
 QY 541 ttgtctgcttctgtctctgtagccctcgtgtaagaatcaatgaggtgagtgatgtgtctgcat 600
 Db 541 TTGCTGCTTCTGTCTCTGTAGCCCTGTGATGAAATCAATGAGGTGATGTTGTCGAT 600
 QY 601 attaccctgacacaaaggaggaacttattcttaccctgcacatctcctatggtttatcatctgtt 660
 Db 601 ATTACCCCTGGCCATAGGGAGACTTATCTTACTCTGACATCTCTCTGTTTATCATTTGTT 660
 QY 661 gctattcccgatccgcacagtagaaggcaaggaggaagggcctctcaacatgctcatc 720
 Db 661 GCTATTCTCCGATCCGACAGTAGAAGAGCAAGAGAGAGAGGCTCTCTCAACTGTCATCT 720
 QY 721 catctcacagtgtagacccttactatctctcctgtaatctaaactatataccgctctgct 780
 Db 721 CATCTCACAGTGTGACCTTACTATTTCTCTGTAATCTACCTATATACGCCCTGCT 780
 QY 781 tccagctatacatcttgaagaagacaagtgtagtgcacacttaactctgtgtactccc 840
 Db 781 TCCAGCTATACATTTGAAAGAGACAAGGTGAGCTGCACACTTATACCTTGTGACTGCC 840
 QY 841 acattaaaccgagatggtgacagcttccagaatagggagatgcaaggaggaattaggag 900
 Db 841 ACATTAACCCGATGSTGTACAGCTTCCAGATAGGGAGATGACAGGAGAAATTAGGAG 900
 QY 901 gttgttgcattcttcgaacac 921
 Db 901 GTTGTGCATTTCTGAAACAC 921

RESULT 5
 AX350633 798 bp DNA linear PAT 06-FEB-2002
 LOCUS AX350633 Sequence 11 from Patent WO01771177.
 DEFINITION AX350633
 ACCESSION AX350633
 VERSION AX350633.1 GI:18616201
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (sites)
 AUTHORS Shenoy,S., Gangoli,I.,E.A., Rastelli,L., Smithson,G., Padigaru,M.,
 Verneet,C.A., Wolner,A.R., Casman,S.J., Tcheney,V.T.,
 Sekeres,E.S., Gorsse,W., Alsobrook,J.P. and Burgess,C.E.
 TITLE Novel spor-proteins and nucleic acids encoding same

JOURNAL Patent: WO 0177177-A 11 18-OCT-2001;
 FEATURES Curagen Corporation (US)
 LOCATION/Qualifiers
 SOURCE 1. 798
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 191 a 202 c 164 g 241 t
 ORIGIN

Query Match 84.3%; Score 778.8; DB 6; Length 798;
 Best Local Similarity 99.7%; Pred. No. 6.2e-211;
 Matches 780; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 143 atagaacaccttggatagcccaatgatagttcttccctctgacactgctgtgtgaca 202
 Db 1 ATACACACCTTGATACGCCCATGATGTTTCTTCTGACACTGGCTGTGTGACA 60
 QY 203 tcatctgacacaaagacatcatcaccgaagatgctgggacatgctaaatcaagaata 262
 Db 61 TCATCTGCACAAACAGCATATACCGAAGATGCTGGGACCATGCTAACATCAGAAAATA 120
 QY 263 ccatttcaatgcaagctgcagctcctcctctctctgttcaaatgctctgggagctg 322
 Db 121 CCATTTCAATATGACAGCTGCATGCTCCAGCTCTTCTTGTTCATAGTCTCTGGAGCTG 180
 QY 323 agatgctctcttccacacacatgacctatgacagctatggtgacattgttccctctc 382
 Db 181 AGATGTTCTCTTCCACACCATGAGCCCTATGACCGCTATGTGGCATTTGTTCCCTTTC 240
 QY 383 attacagtaactatagaaacacatagtgtgtgacctgtctgtctgacatgctatgctga 442
 Db 241 ATTACAGTACTGTTATGAACACACCATATGTGTGATGCTGCTCAGCATGCTATGGCTA 300
 QY 443 ttgacacacaaatctcctgtgtgacacacagctcttatacaggtgagttgactctgtggc 502
 Db 301 TTCAGTACACAAATTCCTGGGTGCACACAGCTCTTATCATGAGGTGACTTCTGTGGCC 360
 QY 503 caaacacacattgacacactctctgtgagatacccccattgtgtgtgtctgtctgtaagc 562
 Db 361 CAACACCAATTTGACCACTTCTGTGAGATACCCCATGTGCTGCTGTGTCTGTAGGCC 420
 QY 563 ctgtaagaatcaatgagtgatgtgtatgtgtctgatacttaccctgacataaggagact 622
 Db 421 CTGTAAAGATCAATGAGGTGATGTTGCTGATATATACCTGCGCATATAGGGGACT 480
 QY 623 ttattcttaccctgacatctcctatgtgtttatactgtgtgtctatctccgtatccgaag 682
 Db 481 TTATTCTTACCTGCATCTCCTATGTGTTTATCATTTGTGCTATTTCTCCGATCCGACAG 540
 QY 683 tagaagacaggaagaaagccttccaaacatgctcatctcatctcaagatggtgacactt 742
 Db 541 TAGAAGCAAGAGAGAGGCTTCTCAACATGCTCATCTCATCTACAGTGTGATACCTT 600
 QY 743 actattctctgtatctatacactatataccgcccctgtctccagctataactttgaagaag 802
 Db 601 ACTATTCTCCTGTATCTATACACTATATCCGCCCTGCTTCCACTATATCAATTTGAAGAG 660
 QY 803 acaagtggtgagctgcacactatactctgtgactccacattaaaccgagatgtgtaca 862
 Db 661 ACAAGTGGTGATGCTGACTCTATACCTCTTGTGACTCCACATTAACCCGATGSTGTACA 720
 QY 863 gcttccagaataggaatgcaaggcaggaataggaaagtgttgcattctctgaacact 922
 Db 721 GCTTCCAAATATAGGAGATGACAGCAGGAATTAGAAGAGTGTGCAATTTCTGAAACACT 780
 QY 923 ag 924
 Db 781 AG 782

RESULT 6
 AC099601/c

LOCUS AC009601 221466 bp DNA linear HTG 16-NOV-2000
DEFINITION Mus musculus clone RP23-240K1, WORKING DRAFT SEQUENCE, 3 unordered pieces.
AC009601.1 GI:16946330
VERSION AC009601.1
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 221466)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE JOURNAL
REFERENCE 2 (bases 1 to 221466)
Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barina, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, Y., Chazaro, B., Chopek, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Gerdes, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamaseras, R., Landers, T., Lehotzky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marguis, N., Matthews, C., McCarthy, M., McEwan, P., McEwan, K., McPheters, R., Meldrum, J., Menusz, L., Mhova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viell, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zemek, L., Zimmer, A. and Zody, M.
DIRECT SUBMISSION
Submitted (16-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A. F. A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
TITLE JOURNAL
COMMENT
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L17240
Center clone name: 240_K-1
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 220831 bases at least Q40
Consensus quality: 221147 bases at least Q30
Consensus quality: 221206 bases at least Q20
Insert size: 230000; agarose-fp
Insert size: 221266; sum-of-contigs
Quality coverage: 11.4 in Q20 bases; agarose-fp
Quality coverage: 11.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated when the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 9911: contig of 9911 bp in length
* 9912 10011: gap of 100 bp

FEATURES	10012	74781	contig of 64770 bp in length
*	74782	74881	gap of 100 bp
*	74882	221466	contig of 146585 bp in length.
SOURCE	1..221466	Location/Qualifiers	
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone="RP23-240K1"		
	/clone_11b="Rpci-23 Female Mouse BAC"		
	1..9911		
misc_feature	/note="assembly_fragment"		
	clone_end:SP6		
	vector_side:left"		
misc_feature	10012..74781		
	/note="assembly_fragment"		
	74882..221466		
misc_feature	/note="assembly_fragment"		
	clone_end:T7		
	vector_side:right"		
BASE COUNT	71567 a	39922 c	39522 g 70247 t 208 others
ORIGIN			
Query Match	74.4%	Score 687.8:	DB 2: Length 221466;
Best Local Similarity	84.1%	Pred No. 9.1e-185;	
Matches 776:	Conservative	0; Mismatches 147;	Indels 0; Gaps 0
QY 1	atgaatcacagcgtgtgtaactgagtcatcattctcgtgcctcaccaaaagcctgaatc	60	
DB 193321	ATGAACCTTCAGCATTCGTATAGTGAGTTTANAGATTCTGCGACTTACTGCAAAAGCTGAACTT	193262	
QY 61	cagggaataatct	120	
DB 193261	CAGGGAAATACCTTTTATCTGTTTTTCTTTTATCTTACCTTGCGCTCTACTGTTAATATG	193202	
QY 121	ctcatcatcatctggcacaatatcatagcaacacctctgcatagcccatgfatgtttccct	180	
DB 193201	CTAATTTGTTGTTCCTAATCTATATACACACCCTTGCACACACCCTATGATATCTTCCTT	193142	
QY 181	ctgaacactgctgtgtgtggaacatcatctgcacaacaagacatcataccgaagatgctg	240	
DB 193141	CTGGCCTTGCGTGTGGTGACAAATCTGCACACAAAGCATCATACCACAAATGTTGGGA	193082	
QY 241	accatgctacaacatcagaataatccatcttatatgacagctgcacatgtccagctctctg	300	
DB 193081	ACTATGTTTACATTCACAAAATTCATTTATATGGGGGTTGGCATGTCACAGCTCTTCTT	193022	
QY 301	tccacatgctctcctggagatgagatggtctctcttcacacacatggccctcagacgcat	360	
DB 193021	TTTCACATGCTCCTTG6GGGCTAGATGAGTGGCTTTTACTACAAATGGCTATAGCCCTAT	192962	
QY 361	gtggccattgtttccctctctcatcacagtaacatattatgaacacacatatgtgtgtagc	420	
DB 192961	GTGGCCATTGTGCTTCCCACTTGGCTATATGTAATGTAATGACACAGATATCAATGTGAGGC	192902	
QY 421	ttgtctcagcatggtcatggtcatctgcagtcaccaatctcctgggtgacacagctctalc	480	
DB 192901	TTGCTTAGCATGTGATGCGCTTGTGCTGTAAACCAATTCCTGGGTCGACACTGGTCTCAT	192842	
QY 481	atgaagttgaccttctgtggtggccaaacacatctgaccactctctctgtgagatacccca	540	
DB 192841	CTGAAGCTGACATTTCTGTGGGCCGAATAAATTAAGACACACTTCTTCTGTGTAATACCCCG	192782	
QY 541	ttgcgtgacctgttctcgtatagccctgtaagaatacaatgaagtgatgtagtattgctgagt	600	
DB 192781	CTGCTTGCTTTTCTCTGATAGTCTGTAAAGTCAATGAGGTGATGATGTTGCTGAT	192722	
QY 601	attaccctggccatagagggaacttattctctactgcatactccatcattgatttcatctgt	660	
DB 192721	ATCACCTGGGCTGTGGGAGACTTTACTGTAACTGCATTCTCTATAGGATTTATTAATGCT	192662	
QY 661	gtcattctccgattccgacagtagaaggcaaggaaaggcctctcaacatgctcatct	720	

Db 192661 GCTATTTCGCCATCCGACACAAGNAGCAAGNAGAGGCGCTTCTCTACGTCCTCATCG 192602
Qy 721 catccacagagggagagcttactatctctctgaatctacacccatctgcctgcct 780
Db 192601 CACCTGACGAGGGGTGCTCCCTTACATTTCTCTCTGATCTACACTATATCCGACCTGCA 192542
Qy 781 tccacatctacatctgaagaagacagaggtgtagctgactctactctactctctgactccc 840
Db 192541 TCCACTGTAACCTTTGATATTAAGACAGAGGTGGTAGCTGCATATATACCTGTAGTACTCT 192482
Qy 841 acattaaacccagatggtgtacagcttccagaatagaggagatgcagagcaggaattaggaag 900
Db 192481 ACATTGACCCCAATGATATATAGCTTCAGGACAAAGAGATGCAATCAGATTTAGAAA 192422
Qy 901 ggtgttgcattctgaacacta 923
Db 192421 GTATTTCGCTTTTAAAGCTTA 192399

RESULT 7
AC094718/c
LOCUS AC094718.2 GI:17941497
DEFINITION Rattus norvegicus clone CH230-5M10, *** SEQUENCING IN PROGRESS ***,
65 unordered pieces.
ACCESSION AC094718
VERSION AC094718.2 GI:17941497
KEYWORDS HTG: HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 160483)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsprouk,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowles,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chenn,C.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Diaper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frazz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsli,F., Howard,S., Huber,J., Huliy,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudh,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korval,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Lounseghed,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwoko,S.,
Ogunu,M., Okwunu,G., Otagunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Peters,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubenkan,I., Rolfe,M.,
Ruliz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshchari,N.,
Slisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Swatek,A., Tabors,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,R.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Washington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Unpublished
2 (bases 1 to 160483)
Worley,K.C.
Direct Submission
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15624554.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GBFP
Center clone name: CH230-5M10
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 127493 bases at least Q40
Consensus quality: 137458 bases at least Q30
Consensus quality: 144882 bases at least Q20
Estimated insert size: 129461; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agatose-fp estimation
Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 65 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
5755 5754: gap of 5754 bp in length
5855 5854: gap of unknown length
11491 11490: contig of 5636 bp in length
11491 11590: gap of unknown length
11591 17216: contig of 5626 bp in length
17217 17316: gap of unknown length
17317 21754: contig of 4438 bp in length
21755 21854: gap of unknown length
21855 25351: contig of 3497 bp in length
25352 25451: gap of unknown length
25452 30551: contig of 5200 bp in length
30552 30751: gap of unknown length
30752 34861: contig of 4110 bp in length
34862 34961: gap of unknown length
34962 38499: contig of 3538 bp in length
38500 38599: gap of unknown length
38600 41784: contig of 3185 bp in length
41785 41884: gap of unknown length
41885 45873: contig of 3969 bp in length
45874 45973: gap of unknown length
45974 50224: contig of 4251 bp in length
50225 50324: gap of unknown length
50325 53266: contig of 2942 bp in length
53267 53366: gap of unknown length
53367 56430: contig of 3064 bp in length
56431 56530: gap of unknown length
56531 59835: contig of 3305 bp in length
59836 59935: gap of unknown length
59936 62863: contig of 2928 bp in length
62864 62963: gap of unknown length
62964 65355: contig of 2392 bp in length
65356 65455: gap of unknown length
65456 68305: contig of 2850 bp in length
68306 68405: gap of unknown length
68406 70717: contig of 2312 bp in length
70718 70817: gap of unknown length
70818 73306: contig of 2489 bp in length

* 73307 73406: gap of unknown length
* 73407 73576: contig of 4170 bp in length
* 77577 77676: gap of unknown length
* 77677 79917: contig of 2241 bp in length
* 79918 80017: gap of unknown length
* 80018 83098: contig of 3081 bp in length
* 83099 83198: gap of unknown length
* 83199 85948: contig of 2750 bp in length
* 85949 86049: gap of unknown length
* 86049 88300: contig of 2252 bp in length
* 88301 88400: gap of unknown length
* 88401 90834: contig of 2434 bp in length
* 90835 90934: gap of unknown length
* 90935 92486: contig of 1552 bp in length
* 92487 92586: gap of unknown length
* 92587 94928: contig of 2342 bp in length
* 94929 95028: gap of unknown length
* 95029 97296: contig of 2268 bp in length
* 97297 97396: gap of unknown length
* 97397 99357: contig of 1961 bp in length
* 99358 99457: gap of unknown length
* 99458 101633: contig of 2176 bp in length
* 101634 101733: gap of unknown length
* 101734 103947: contig of 2214 bp in length
* 103948 104047: gap of unknown length
* 104048 106350: contig of 2303 bp in length
* 106351 106450: gap of unknown length
* 106451 108776: contig of 2326 bp in length
* 108777 108876: gap of unknown length
* 108877 111586: contig of 2710 bp in length
* 111587 111686: gap of unknown length
* 111687 113940: contig of 2254 bp in length
* 113941 114040: gap of unknown length
* 114041 115113: contig of 1073 bp in length
* 115114 115213: gap of unknown length
* 115214 116454: contig of 1241 bp in length
* 116455 116554: gap of unknown length
* 116555 118145: contig of 1591 bp in length
* 118146 118245: gap of unknown length
* 118246 119842: contig of 1597 bp in length
* 119843 119942: gap of unknown length
* 119943 121213: contig of 1271 bp in length
* 121214 121313: gap of unknown length
* 121314 123957: contig of 2644 bp in length
* 123958 124057: gap of unknown length
* 124058 125239: contig of 1182 bp in length
* 125240 125339: gap of unknown length
* 125340 126844: contig of 1505 bp in length
* 126845 126944: gap of unknown length
* 126945 128704: contig of 1760 bp in length
* 128705 128804: gap of unknown length
* 128805 130980: contig of 2176 bp in length
* 130981 131080: gap of unknown length
* 131081 132857: contig of 1777 bp in length
* 132858 132957: gap of unknown length
* 132959 134901: contig of 1944 bp in length
* 134902 135001: gap of unknown length
* 135002 136075: contig of 1074 bp in length
* 136076 136175: gap of unknown length
* 136176 137403: contig of 1228 bp in length
* 137404 137503: gap of unknown length
* 137504 139120: contig of 1617 bp in length
* 139121 139220: gap of unknown length
* 139221 141101: contig of 1881 bp in length
* 141102 141201: gap of unknown length
* 141202 143536: contig of 2335 bp in length
* 143537 143636: gap of unknown length
* 143637 144839: contig of 1203 bp in length
* 144840 144939: gap of unknown length
* 144940 146098: contig of 1159 bp in length
* 146099 146198: gap of unknown length
* 146199 147377: contig of 1179 bp in length
* 147378 147477: gap of unknown length

* 147478 148718: contig of 1241 bp in length
* 148719 148818: gap of unknown length
Query Match 72.1%; Score 666.6; DB 2; Length 160483;
Best Local Similarity 82.6%; Pred. No. 9,8e-179;
Matches 762; Conservative 0; Mismatches 161; Indels 0; Gaps 0;
QY 1 atgaatcacagcgttgtaactgaatcattacattcggcgtccacaaagcctgaactc 60
DB 13579 atgaactttagcattgtttatagatttttgaggacttaccataaactgactt 13520
QY 61 cagggaatcattcctccttttctcattgtcattcctgtggtcttcctggcacatg 120
DB 13519 cagggaatcattcctccttttctcattgtcattcctgtggtcttcctggcacatg 13460
QY 121 cccatcatcattgccaaaatcattagcaacaccttgcaagcccatgtaattcctt 180
DB 13459 nttaattgttggtccattatttataacaccacttgccacacacattgtatttctctt 13400
QY 181 ctgacacgtgctgtgtgacatcatctgcacaaacagatcatccagagatgtcgtg 240
DB 13399 ctggcccttgctgtgtgacatcatctgcacaaacagatcatccagagatgtcgtg 13340
QY 241 accatgctaacatcagaataacatttcaatattgagagctgcatgtccagctctctg 300
DB 13339 acatgtttacattcagaaataatgcatttcatatgagggcgctgcatgcccagctctctt 13280
QY 301 ttcacatgctctcctggagctgagatgtctccttccacacacatggtcctatgc 360
DB 13279 tttaactgttcttttggggcctgagatgtgcttcttccatcaccatggtcctatgc 13220
QY 361 gtggcattgttctccttccattacagttactattatgaacacattggtgtgagcc 420
DB 13219 gtggcattgttctccttccattacagttactattatgaacacattggtgtgagcc 13160
QY 421 ttgctcagatctgtaattggtcatttcgaatcacaatcttcctgggtgacacagcttctc 480
DB 13159 ttgctcagatctgtaattggtcatttcgaatcacaatcttcctgggtgacacagcttctc 13100
QY 481 atgaggttgaacttctctgtggtgccaacacacattgcacacttctctgtgagatcccca 540
DB 13099 cttaagcctgacttctctgtggtgccaacacacattgcacacttctctgtgagatcccca 13040
QY 541 ttgctgacttctctgtggtgccaacacacattgcacacttctctgtgagatcccca 600
DB 13039 ctgctgtgcttctctgtggtgccaacacacattgcacacttctctgtgagatcccca 12980
QY 601 attaccctggccatagagggaacttattcttaccatcattcctcattgatttcaatctt 660
DB 12979 attaccctggccatagagggaacttattcttaccatcattcctcattgatttcaatctt 12920
QY 661 gctattctcgtatccgcacagtagaaggcaaggagccttctcacaatgctcattc 720
DB 12919 gctattctcgtatccgcacagtagaaggcaaggagccttctcacaatgctcattc 12860
QY 721 catcccaagtgtagaccttactatctccttgaatcattacattatccgcctctc 780
DB 12859 catcccaagtgtagaccttactatctccttgaatcattacattatccgcctctc 12800
QY 781 tccaactatatacttgaaagagacaaggtygtagctgacattacatttctgactcc 840
DB 12799 tccaactatatacttgaaagagacaaggtygtagctgacattacatttctgactcc 12740
QY 841 acattaaaccgaatggtgtacagcttccagaatagagagatgcaggcgagaaatggaag 900
DB 12739 acattaaaccgaatggtgtacagcttccagaatagagagatgcaggcgagaaatggaag 12680
QY 901 gttgttgcattctgaacacta 923
DB 12679 gttgttgcattctgaacacta 12657
RESULT 8

AY073900 926 bp DNA linear ROD 04-FEB-2002
 LOCUS AY073900 Mus musculus olfactory receptor MOR251-4p pseudogene, partial
 DEFINITION Mus musculus olfactory receptor MOR251-4p pseudogene, partial
 sequence.
 AY073900
 ACCESSION AY073900.1 GI:18481062
 VERSION
 KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 926)
 AUTHORS Zhang, X. and Firestein, S.
 TITLE The olfactory receptor gene superfamily of the mouse
 JOURNAL Nat. Neurosci. 5 (2), 124-133 (2002)
 PUBMED 11802173
 REFERENCE 2 (bases 1 to 926)
 AUTHORS Adams, M.
 TITLE Direct Submission
 JOURNAL Submitted (11-JAN-2002) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 FEATURES
 source
 1. 926
 Location/Qualifiers
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="7"
 <1..>926
 /gene="olfactory receptor MOR251-4p"
 /note="potential coding region disrupted by one or more
 internal stop codons"
 /pseudo
 /evidence=not-experimental
 BASE COUNT 227 a 209 c 181 g 309 t
 ORIGIN
 Query Match 72.0%; Score 665.6; DB 10; Length 926;
 Best Local Similarity 84.1%; Pred. No. 1.1e-178;
 Matches 774; Conservative 0; Mismatches 144; Indels 2; Gaps 2;

Db 484 CTGAGGCTGACTTCTGTGGCGCAATATATGACACATCTTCTGGAATACCC 543
 Qy 539 cattgctgcttgcctgtacgcctgaagaatcaatgagtgatgctgctg 598
 Db 544 CGTCTGTGCTTGTCTGCTGAGTCTGTAAGAGCAATGAGTGTATGCTGCG 603
 Qy 599 atattaccctggacatagggaacttattcttactgcatctcctatggtttatcatg 658
 Db 604 ATATACCCCTGCTGTGGAGACTTACTCAACCTCATCTTCTATGATTTATTTG 663
 Qy 659 ttgtattcttcgcatccgcagtagaaggcaaggagggctcttcaactgcat 718
 Db 664 CTGCATATCTCCGCATCCGCAACAAGAGCAAGAGAGAGGCTTCTCACTGCTCAT 723
 Qy 719 ctcaatcagaagtgtagaccccttactatcttcctgtaactatcaactatccgcctg 778
 Db 724 CGACCTCATAGTGTGCTGCTTACTATCTCTGTTATCTACCATATCCGACCTG 783
 Qy 779 ctccagctatacatltgaaagagaagaagtgtagctgcacctatactctgtgactc 838
 Db 784 CATCCAGCTATACCTTGTATTAAGACAAGGTGTAGCTGCATATATCTTACTTAATC 843
 Qy 839 ccacattaaacccgtagtgtagacagctccagaaatagagatgcaggcaggaaataga 898
 Db 844 CTACATGTGAACCAAT 903
 Qy 899 aggtgttgcatcttctgaa 918
 Db 904 AAGTATTTGCATTTTAA 923
 RESULT 9
 AX350675 663 bp DNA linear PAT 06-FEB-2002
 LOCUS AX350675 Sequence 53 from Patent WO0171717.
 DEFINITION AX350675
 ACCESSION AX350675
 VERSION AX350675.1 GI:18616221
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (sites)
 AUTHORS Shenoy, S., Gangolli, E.A., Rastelli, L., Smithson, G., Padigaru, M.,
 Vernet, C.A., Wolenc, A.R., Casman, S.J., Tcheney, V.T.,
 Szekeres, E.S., Gorsse, W., Alsobrook, J.P., and Burgess, C.E.
 TITLE Novel opcr proteins and nucleic acids encoding same
 JOURNAL Patent: WO 0171717-A 53 18-OCT-2001;
 Curagen Corporation (US)
 FEATURES
 source
 1. 663
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 151 a 174 c 134 g 204 t
 ORIGIN
 Query Match 71.6%; Score 661.4; DB 6; Length 663;
 Best Local Similarity 99.8%; Pred. No. 1.7e-177;
 Matches 662; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY	367	attcttttccccccttcataacagactatctatgaaccaccatatgtgtaagccttgcc	426
Db	181	ATTGTGTTTCCTCCTTCATTAACAGTACTGTTAAGAACCACCATATGtGTATGCCTTGCTC	240
OY	427	agcatgycatcgctatgcaatgcagtcaccaatlccctgggtgcacacagcctcttalcagag	486
Db	241	AGCATGTCATGGCGTATTATTCAGTCAACCAATTCCTGGGTGCACACAAGCCTTTATTCATGAGG	300
OY	487	ttagcttctcgtggtggccaacaaccatgtgcaaccttcctcgtgagaataccccatgtcg	546
Db	301	TTAGCTTTCTGTGGGCCAAACACCAATTGACCACTTCCTCTGTGAATATCCCCCAATTGGCTG	360
OY	547	gctttgttcctgtaaagcccttgaagaatcaaaggagtgatggtgatgttgttatgttatctacc	606
Db	361	GCTTTGTCCTGTGAAGCCCTGTAAAGTAACAATGAGTGTATGTATGTTGCTATATTTAAC	420
OY	607	ctggccataaggaggaatttattcttaccgtgcactccctaagtttatacatigtctatt	666
Db	421	CTGGCCATAAGGGGACTTTATTTCTTAOCGCAATCCTCATNGTTTATATATTTTGCTAT	480
OY	667	cctcgatccgcacagtagaagaagaaggaccccttcacaatgtcatctcatctc	726
Db	481	CCTCCCTATCCGCACAGTAGAAGCAAGGAAGGAGCGCTTCTCAACATGCTCATCATCTGC	540
OY	727	aagagtgtagaccccttactatcttcctcgttaactbaactatatacgcgcctgtccagc	786
Db	541	AAAGTGAGGAGCCCTTACTATTCTCTCTGTAATCTACACCTATATTCGCCCTGCTTCAGC	600
OY	787	tatacatcttgaagaagaaggtagtgtagctgcactctaatctctgtgactccacatla	846
Db	601	TATACATTGGAAGAAGACAGGTGTAGCTGCACACTATACCTTGTTAGCTCCACATTA	660
OY	847	aac 849	
Db	661	AAC 663	
RESULT 10			
LOCUS	AC101272	59688 bp	DNA linear HTG 23-NOV-2001
DEFINITION	Mus musculus clone RP23-10J20, LOW-PASS SEQUENCE SAMPLING.		
ACCESSION	AC101272		
VERSION	AC101272.1 GI:17060047		
KEYWORDS	HTG; HTGS; PHASEO.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Euhayrrola, Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 59688)		
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.		
TITLE	Mus musculus, clone RP23-10J20		
JOURNAL	unpublished		
REFERENCE	2 (bases 1 to 59688)		
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Bartha,N., Bastien,V., Boguslavskiy,L., Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,Y.J., Chazarro,B., Chapel,V., Colangelo,M., Collins,K., Collamore,A., Cook,A., Cooke,P., Derrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Grinde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Margus,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McNetters,R., Meldrum,J., Menes,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre-N., Pollara,V., Raymond,C., Retta,R., Rieback,N., Riley,R., Rise,C., Rogov,P., Roman,J., Roselli,M., Roy A., Santos,R., Schauer,S., Schnpack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,		

TITLE
JOURNAL
COMMENT

Topham, K., Travers, M., Travs, N., Trifillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wymn, D., Ye, W.J., Young, C.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center -----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www.seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L16346
Center clone name: 10L_C_20

* NOTE: This record contains 76 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```
1
* 680 779: contig of 679 bp in length
*      100 bp
* 780 1489: contig of 710 bp in length
* 1490 1589: gap of 100 bp
* 1590 2263: contig of 674 bp in length
* 2264 2363: gap of 100 bp
* 2364 3054: contig of 691 bp in length
* 3055 3154: gap of 100 bp
* 3155 3839: contig of 685 bp in length
* 3840 3939: gap of 100 bp
* 3940 4630: contig of 691 bp in length
* 4631 4730: gap of 100 bp
* 4731 5401: contig of 671 bp in length
* 5402 5501: gap of 100 bp
* 5502 6164: contig of 663 bp in length
* 6165 6264: gap of 100 bp
* 6265 6949: contig of 685 bp in length
* 6950 7049: gap of 100 bp
* 7050 7728: contig of 679 bp in length
* 7729 7828: gap of 100 bp
* 7829 8516: contig of 688 bp in length
* 8517 8616: gap of 100 bp
* 8617 9325: contig of 709 bp in length
* 9326 9425: gap of 100 bp
* 9426 10127: contig of 702 bp in length
* 10128 10227: gap of 100 bp
* 10228 10924: contig of 697 bp in length
* 10925 11024: gap of 100 bp
* 11025 11712: contig of 688 bp in length
* 11713 11812: gap of 100 bp
* 11813 12441: contig of 669 bp in length
* 12482 12581: gap of 100 bp
* 12582 13254: contig of 673 bp in length
* 13255 13354: gap of 100 bp
* 13355 14033: contig of 679 bp in length
* 14034 14133: gap of 100 bp
* 14134 14816: contig of 683 bp in length
* 14817 14916: gap of 100 bp
* 14917 15597: contig of 681 bp in length
* 15598 15697: gap of 100 bp
* 15698 16388: contig of 691 bp in length
* 16389 16488: gap of 100 bp
* 16489 17234: contig of 746 bp in length
* 17235 17334: gap of 100 bp
```


BASE COUNT	185 a	279 c	198 g	253 t
ORIGIN	ASIIKIRISABEGKORASTCASHLLIVIMYYTYITIRPSSSYLKKDKVVISLIYNS VAPPLNPLIYTLNRKDKVALRLFSCC			

Query Match	40.0%	Score 369.2	DB 10	Length 915
Best Local Similarity	63.6%	Pred. No. 3,4e-94		
Matches 579	Conservative	0	Mismatches 320	Indels 3
Gaps				
1	atgaatacaaggtgtgtactgaatcattatcttctggcctcaacaaagcctgaac	60		
1	ATGAAATGAGGACACTGCTACTGATGTCCATCTCCAGTTCGATATGCTCACTT	60		
61	cagggaatatct	120		
61	CGGAATCACTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	120		
121	ctccatcatcttgcgaataatctatgaacacacttgcacagcccatgtatgtt	180		
121	CTCATTTATGCTGACATGATGTCGACCCCAACCCCTCATACCCCTTGTACTCTTCC	180		
181	ctgaacatgctgt	240		
181	GTCACACTGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	240		
241	accatgctcaacatcaagaataatcattatcattatgcagctgtcatgttcc	300		
241	ACCATGT	300		
301	ttcaaatgtctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	360		
301	TTCACATGCTGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	360		
361	gtggcatttgttctctctctctctctctctctctctctctctctctctct	420		
361	GTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	420		
421	ttgtctcaagatgtcatgtctatgtcagttacccaattccgggtgtgtgtgt	480		
421	TTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	480		
481	atgaatgtgacttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	540		
481	ATGAAATGAGGACACTGCTACTGATGTCCATCTCCAGTTCGATATGCTCACTT	540		
541	ttgtctgtcttctgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	600		
541	TTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	600		
601	attacacatgt	660		
601	ATGAAATGAGGACACTGCTACTGATGTCCATCTCCAGTTCGATATGCTCACTT	660		
661	gtcatttctcgtatccgcaagaatgaagaaggaagaagaaggaagaaggaag	720		
661	GTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	720		
721	catctcaagtggt	780		
721	CATCTCAAGT	780		
781	tcgaatataatcttgaagaagaagaagtgtagtgaactcttaactctgtgtcc	840		
781	TCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	840		
841	acataaaccgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	900		
841	ACATTAACCGAATGT	900		
901	gt	960		
901	GTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	960		
998	CTTATCT	1056		
998	CTTATCT	1056		

RESULT 13
AC094870/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE

REFERENCE

COMMENT

AC094870 112730 bp DNA linear HTG 20-DEC-2001
Rattus norvegicus clone CH230-5N18, *** SEQUENCING IN PROGRESS ***
52 unordered pieces.
AC094870
AC094870.2 GI:17941654
HTG: HTGS.PHASE1.
Norway rat.
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 112730)
Muzny,D.M., Adams,C., Adio-Ogunola,B., Ali-osman,F.R., Allen,C.,
Alshrooks,S.L., Amaralungs,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimge,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Briera,M., Brown,E., Brown,M., Bryant,N.P., Bulhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleaveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinu,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Dublin,K.J., Eamhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisl,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Huiyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kunesht,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Lounsgesed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., Mcleod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokweto,S.,
Ogih,M., Okumura,G., Ogunyeye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojuboakan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,
Slitson,I., Sodergren,E., Sonalke,J., Sparks,A., Stanley,H.,
Stang,H., Sutton,A., Svatek,A., Tabot,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Washington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 112730)
Worley,K.C.
Direct Submission
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15624706.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
Project information
Center project name: GRPC
Center clone name: CH230-5N18

Summary Statistics

Assembly program: Phrap; version 0.990329First call to findphraplist

Consensus quality: 96697 bases at least Q40
 Consensus quality: 104847 bases at least Q30
 Consensus quality: 112341 bases at least Q20
 Estimated insert size: 92352; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 1.1x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently consists of 52 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 6916: contig of 6916 bp in length
 6917 7016: gap of unknown length
 7017 11089: contig of 4073 bp in length
 11090 11189: gap of unknown length
 11190 14517: contig of 3328 bp in length
 14518 14617: gap of unknown length
 14618 17992: contig of 3375 bp in length
 17993 18092: gap of unknown length
 18093 21658: contig of 3566 bp in length
 21659 21758: gap of unknown length
 21759 25307: contig of 3549 bp in length
 25308 25407: gap of unknown length
 25408 27818: contig of 2411 bp in length
 27819 31526: contig of 3608 bp in length
 31527 31626: gap of unknown length
 31627 33967: contig of 2341 bp in length
 33968 34067: gap of unknown length
 34068 36742: contig of 2575 bp in length
 36743 39205: contig of 2463 bp in length
 39206 39305: gap of unknown length
 39306 41281: contig of 1976 bp in length
 41282 41381: gap of unknown length
 41382 45065: contig of 3684 bp in length
 45066 45165: gap of unknown length
 45166 47143: contig of 1978 bp in length
 47144 47243: gap of unknown length
 47244 48499: contig of 1256 bp in length
 48500 50055: contig of 1456 bp in length
 50056 50155: gap of unknown length
 50156 53530: contig of 3375 bp in length
 53531 53630: gap of unknown length
 53631 55445: contig of 1815 bp in length
 55446 55545: gap of unknown length
 55546 58015: contig of 2470 bp in length
 58016 58115: gap of unknown length
 58116 59612: contig of 1497 bp in length
 59613 59712: gap of unknown length
 59713 61702: contig of 1990 bp in length
 61703 61802: gap of unknown length
 61803 63157: contig of 1355 bp in length
 63158 63257: gap of unknown length
 63258 64623: contig of 1366 bp in length
 64624 64723: gap of unknown length
 64724 65766: contig of 1043 bp in length
 65767 65866: gap of unknown length
 65867 67869: contig of 2003 bp in length
 67870 67969: gap of unknown length
 67970 68560: contig of 1591 bp in length
 68561 69660: gap of unknown length
 69661 71980: contig of 2320 bp in length
 71981 72080: gap of unknown length

72081 73657: contig of 1577 bp in length
 73658 73757: gap of unknown length
 73758 75695: contig of 1938 bp in length
 75696 75795: gap of unknown length
 75796 78395: contig of 2600 bp in length
 78396 78495: gap of unknown length
 78496 79707: contig of 1212 bp in length
 79708 79807: gap of unknown length
 79808 81374: contig of 1567 bp in length
 81375 81474: gap of unknown length
 81475 83163: contig of 1689 bp in length
 83164 83263: gap of unknown length
 83264 85474: contig of 2211 bp in length
 85475 85574: gap of unknown length
 85575 87264: contig of 1650 bp in length
 87265 87364: gap of unknown length
 87365 88370: contig of 1006 bp in length
 88371 88470: gap of unknown length
 88471 89719: contig of 1249 bp in length
 89720 89819: gap of unknown length
 89820 91531: contig of 1712 bp in length
 91532 91631: gap of unknown length
 91632 92853: contig of 1222 bp in length
 92854 92953: gap of unknown length
 92954 95061: contig of 2108 bp in length
 95062 95161: gap of unknown length
 95162 97151: contig of 1990 bp in length
 97152 97251: gap of unknown length
 97252 98627: contig of 1576 bp in length
 98628 98927: gap of unknown length
 98928 100300: contig of 1373 bp in length
 100301 100400: gap of unknown length
 100400 101467: contig of 1067 bp in length
 101468 101567: gap of unknown length
 101568 102772: contig of 1205 bp in length
 102773 102872: gap of unknown length
 102872 104128: contig of 1256 bp in length
 104129 104228: gap of unknown length
 104229 105514: contig of 1286 bp in length
 105515 105614: gap of unknown length
 105615 106797: contig of 1183 bp in length
 106798 106897: gap of unknown length
 106898 108212: contig of 1315 bp in length
 108213 108312: gap of unknown length
 108313 109627: contig of 1315 bp in length
 109628 109727: gap of unknown length
 109728 110921: contig of 1194 bp in length
 110922 111021: gap of unknown length
 111022 112730: contig of 1709 bp in length.

FEATURES
 source 1. 112730
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-SN18"

BASE COUNT 30919 a 22306 c 22742 g 31565 t 5198 others
 ORIGIN

Query Match 36.7%; Score 339; DB 2; Length 112730;
 Best Local Similarity 62.6%; Pred. No. 2.2e-85;
 Matches 563; Conservative 0; Mismatches 330; Indels 6; Gaps 2;

QY 1 atgaatcaagaggttgtaactgaagttacattatcttggcgtccacaaagctgaactc 60
 DB 96341 ATGAATGGGACACTGGTCTCTAGTTCCTCAATCCGAGGATTCAGAAATGCTCACCTT 96282
 QY 61 cagggaattatcttcctcttttctcatctgtcattctctgtgctttctcgcaaatg 120
 DB 96281 CGGG---TACCACTTCTCAGCTTCCTCCTTCATATCATGCGTCAGCAATCTCAGGAACCTG 96225
 QY 121 ctcatcatcatgtgcaaaaatctatagcaaaccttgcatagcaccatgtaigtcttctt 180
 DB 96224 CTCATATGCTGACAGTCAGTCCAGCCAGCCCTGACATACCCCTATGACTTCTCTG 96165

QY	181	ctgcacatgctgtttgtggacatcatctgtgacaacaagccttaacgaagatgctggg	240
Db	96164	GTCAACTTAACTGTGTGGATGACATCTCTGGACCTCCACCACTCACTTCCCAAGCTCTGGAC	96105
QY	241	accatgtcaacatcagaanaatcaacatcttcaatgatgagctgtgcatgtgtccagcttcttg	300
Db	96104	ATCANGT---AGGGGGAGAGACCATCTCTATGTGGTGTGCATAGGCCACGCTCTTCTTC	96048
QY	301	ttcacatggtctctgggagctgtgatatgttcttcaacacacatgagctcttgacagctat	360
Db	96047	TTCACATGGTGTCCATGGGGGAGAGCTTGTGTCTCTCAAGTAAGGCTTATGACCGCTTC	95988
QY	361	gtggcgcatttcttcccttcaatcaagatactatlgaaaccaacatlatgtgtgtgacc	420
Db	95987	GTGGCAATCTGCTGTGCCCTGTGCATATATGTACGTGGATGGGCCCAAGGATGTGTGCATTC	95928
QY	421	ttgtctcagatgtgcatgtgctattgtgagtcacaaattctctgtgtgcacacagctctatc	480
Db	95927	CTGGCTGGCGCTTGTGTGGTCCATCGACCTGTGACTAACACCACTGTGAACAAGCTGTGGTG	95868
QY	481	atgaggttgaccttctgtggggccaacaacatltgacaactctctctgtgagatacccca	540
Db	95867	CTGCATCTAACCACTCTGCAGCTCCAAATGTGATAGAGCACTTCTGTGTGAAATTCGCCCA	95808
QY	541	ttgtcgtcttctgtcctgtgagccctgtlaagaatcaatgtaggtgtatgtgttgcgtgat	600
Db	95807	CGGTTAAAGCTCTTGTGTCTCCAAACAAATGGAAGGATTTAGGCGCTTGTGTGCATAT	95748
QY	601	attaccctgtcagatggaggaacttatcttctactgcattctcctatgagttatcaatgtt	660
Db	95747	GTGTTTCTGGCTGTAGGGAATCTTCTGTGTGATCATCTCTCTATATGGCTTATATGTGACC	95688
QY	661	gcatcttcgcgtatccgcacagtagaaggtcaagaaggaagccctctcaaatgtcatct	720
Db	95687	AGTATCTGTGAAGTATAGCTCAACTGAGGGCAAGCAAGCGCTTTTCTACCTGTCTCTCA	95628
QY	721	catctcaagttgtgacccttactatcttctctgtaatctatacctatactccgcctgt	780
Db	95627	CATCTCTTGTGTGCACCAATGATATCTCTCTCATATCATCTATACCTACATTTGGCCTGA	95568
QY	781	ttcagctatatactltgaagagagacaagttgttagctgtcaactctatactcttgtgactcc	840
Db	95567	TTCAAGCTACTCATTTGAACAAGACAGAGGTGTGTCCATCATCTACACTTCGGTGGCACC	95508
QY	841	acattaaaccgatgtgtgatacgtccaaagttaaggagagatgaggtcaggtcaggaattgaa	899
Db	95507	ACCTTGAACCCCTCATCTAACCTCTGTAGGAACAAGAATGTCAAAATGTCAGCTCCGGAA	95449

RESULT 14	AC094700/c	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
	AC094700	197438 bp	DNA linear	HTG 20-DEC-2001				
	Rattus norvegicus clone CH230-5F7,	***	SEQUENCING IN PROGRESS ***					
	96 unordered pieces.							
	AC094700							
	AC094700.4	GI:17941479						
	HTG; HTGS-PHASE1.							
	Norway rat.							
	Rattus norvegicus							
	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;							
	Rattus							
	1 (bases 1 to 197438)							
REFERENCE								
AUTHORS								
	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,							
	Alsbrooks,S.L., Amaralunga,H.C., Aye,J.R., Banks,T., Barbieri,J.,							
	Benton,C., Bimge,K., Blankenburg,K., Bonnin,D., Bouck,J.J.,							
	Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,							
	Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,							
	Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,							
	Chen,Z., Choudhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,							
	Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,							
	Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,							

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elahi,C., Escotto,M., Falls,T., Ferraguto,C., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
Hernandez,U., Hernandez,O., Hodson,A., Hognes,M., Holloway,C.,
Hollins,B., Homsi,P., Howard,S., Huber,J., Huly,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kurehli,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtearge,O., Lien,C., Liu,J., Liu,W.,
Lounsbeged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawlinney,E., Mcleod,M.P., Meador,M.,
Mei,G., Metker,M., Miner,C., Miner-Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenwo,S.,
Ogden,M., Okunou,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherter,S., Scott,G., Shen,H., Shoshitari,N.,
Russon,I., Sodegren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Swack,A., Tabor,P., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,F., Telford,B., Thomas,K.,
Thomas,S., Usmali,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Wallington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 197438)

Worley,K.C.
Direct Submission
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:117062158.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----
Center project name: GBEM
Center clone name: CH230-5F7
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329first call to
findphraplist

Consensus quality: 163315 bases at least Q40
Consensus quality: 177813 bases at least Q30
Consensus quality: 189359 bases at least Q20
Estimated insert size: 163704, sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 1.8x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)
* NOTE: This is a "working draft" sequence. It currently
* consists of 96 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1	6380: contig of 6380 bp in length	
*	6381	6480: gap of unknown length
*	6481	11235: contig of 4755 bp in length
*	11236	11335: gap of unknown length
*	11336	14082: contig of 2747 bp in length
*	14083	14182: gap of unknown length

```

* 14183 18911: contig of 4729 bp in length
* 18912 19011: gap of unknown length
* 19012 22084: contig of 3073 bp in length
* 22085 22184: gap of unknown length
* 22185 26199: contig of 4015 bp in length
* 26200 26299: gap of unknown length
* 26300 28568: contig of 2269 bp in length
* 28569 28669: gap of unknown length
* 28669 31639: contig of 2971 bp in length
* 31640 31739: gap of unknown length
* 31740 33776: contig of 2037 bp in length
* 33777 33876: gap of unknown length
* 33877 37642: contig of 3766 bp in length
* 37643 37743: gap of unknown length
* 37743 39938: contig of 2195 bp in length
* 39938 40037: gap of unknown length
* 40038 43330: contig of 3293 bp in length
* 43331 43430: gap of unknown length
* 43431 46413: contig of 2983 bp in length
* 46414 46513: gap of unknown length
* 46514 49924: contig of 3411 bp in length
* 49925 50024: gap of unknown length
* 50025 52777: contig of 2753 bp in length
* 52778 52877: gap of unknown length
* 52878 55322: contig of 2445 bp in length
* 55323 55422: gap of unknown length
* 55423 57118: contig of 1696 bp in length
* 57119 57218: gap of unknown length
* 57219 59661: contig of 2443 bp in length
* 59662 59761: gap of unknown length
* 59762 62279: contig of 2518 bp in length
* 62280 62379: gap of unknown length
* 62380 65763: contig of 3384 bp in length
* 65764 65863: gap of unknown length
* 65864 68840: contig of 2977 bp in length
* 68841 70763: contig of 1823 bp in length
* 70764 70863: gap of unknown length
* 70864 73785: contig of 2922 bp in length
* 73786 73885: gap of unknown length
* 73886 76368: contig of 2483 bp in length
* 76369 76468: gap of unknown length
* 76469 79049: contig of 2581 bp in length
* 79050 79149: gap of unknown length
* 79150 82088: contig of 2939 bp in length
* 82089 82188: gap of unknown length
* 82189 84182: contig of 1994 bp in length
* 84183 84282: gap of unknown length
* 84283 87125: contig of 2843 bp in length
* 87126 87225: gap of unknown length
* 87226 88998: contig of 1773 bp in length
* 88999 89098: gap of unknown length
* 89099 92015: contig of 2917 bp in length
* 92016 92115: gap of unknown length
* 92116 94512: contig of 2397 bp in length
* 94513 94612: gap of unknown length
* 94613 96876: contig of 2264 bp in length
* 96877 96976: gap of unknown length
* 96977 98801: contig of 1825 bp in length
* 98802 100728: contig of 1827 bp in length
* 100729 100828: gap of unknown length
* 100829 102598: contig of 1770 bp in length
* 102599 102698: gap of unknown length
* 102699 104352: contig of 1554 bp in length
* 104353 104352: gap of unknown length
* 104353 105818: contig of 1466 bp in length
* 105819 105918: gap of unknown length
* 105919 107016: contig of 1098 bp in length
* 107017 107116: gap of unknown length
* 107117 108700: contig of 1584 bp in length
* 108701 108800: gap of unknown length
* 108801 110130: contig of 1330 bp in length

```

```

* 110131 110230: gap of unknown length
* 110231 111400: contig of 1170 bp in length
* 111401 111500: gap of unknown length
* 111501 113819: contig of 2319 bp in length
* 113820 113919: gap of unknown length
* 113920 115967: contig of 2048 bp in length
* 115968 116067: gap of unknown length
* 116068 117164: contig of 1097 bp in length
* 117165 117264: gap of unknown length
* 117265 118813: contig of 1549 bp in length
* 118814 118913: gap of unknown length
* 118914 120583: contig of 1670 bp in length
* 120584 120683: gap of unknown length
* 120684 122741: contig of 2058 bp in length
* 122742 122841: gap of unknown length
* 122842 125104: contig of 2263 bp in length
* 125105 125204: gap of unknown length
* 125205 127018: contig of 1814 bp in length
* 127019 127118: gap of unknown length
* 127119 128590: contig of 1472 bp in length
* 128591 128690: gap of unknown length
* 128691 129819: contig of 1129 bp in length
* 129820 129919: gap of unknown length
* 129920 131056: contig of 1137 bp in length
* 131057 131156: gap of unknown length
* 131157 133124: contig of 1968 bp in length
* 133125 133224: gap of unknown length
* 133225 135098: contig of 1874 bp in length
* 135099 135198: gap of unknown length
* 135199 136133: contig of 1415 bp in length
* 136140 136713: gap of unknown length
* 136714 138549: contig of 1856 bp in length
* 138550 138649: gap of unknown length

Query Match      36.7%; Score 339; DB 2; Length 197438;
Best Local Similarity 62.6%; Pred. No. 2,3e-85;
Matches 563; Conservative 0; Mismatches 330; Indels 6; Gaps 2;

QY 1 atgaatcacagcgttgtaactagttcattatctctggtccaccacaaagcctgaactc 60
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 96743 ATGAATGAGGACATGCTGCTACGTGCTTCATCTGAGATTCAGAAATGCTCACCTT 96684

QY 61 cagggaatatctctctctctctctctctctctctctctctctctctctctctctctct 120
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 96683 CGGG--TACCACTTCTCAGCTTCTCTCTCTATCATGCTGCAATTCAGAAACCTTG 96627

QY 121 ctcatcatcatctgccaatactatagcaacacctgcatagcccatatgtcttctct 180
    ||||| | | | | | | | | | | | | | | | | | | | | | | | |
DB 96626 CTCATTATGTTGATGACGTGACGTGACATCTCTGACCTCCACCATCTCAAGCTCTGGAC 96567

QY 181 ctgacactgctgtgtgtgacatcatctgcacacaaagcatcataccgaagatcgtggg 240
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 96566 GTCACTTGTGCTGTGTGACATCTCTGACCTCCACCATCTCAAGCTCTGGAC 96507

QY 241 accatgctacactgaataatcaccttcatatgtagcaggtgatgtccagctctcttg 300
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 96506 ATCATGCT--AGGGGAGAGACCATCTTATGAGTGCTGATGGCCCAAGCTCTTCTTC 96450

QY 301 ttacatgctctctggaagctgagatggtctctctacacacatgagccctatagccgtat 360
    ||||| | | | | | | | | | | | | | | | | | | | | | | |
DB 96449 TTCACATGCTCATGAGGGGACAGAGCTTCGCTTCTGACATGACCTTATGACCGCTTC 96390

QY 361 gtggccattgttccctcttcaatcagactactatatagaaccacacatatgtgtagcc 420
    ||||| | | | | | | | | | | | | | | | | | | | | | | |
DB 96389 GTGGCAATCTGCTGCCCTGACATATATGATACCTGATGGGCCCAAGGCTGTGCTATTC 96330

QY 421 ttgtcagagatgtagtctatctgacagcacaactctgtgtgacacagctctatc 480
    || | | | | | | | | | | | | | | | | | | | | | | | |
DB 96329 CTGGCTGCTGTGTGCTGCTCAATCAGCTGATTAACCAAGTGTACCAAGCCTGCTG 96270

QY 481 atgaggttgaactctgtgtgccaacacacatgacacatctctctgtgtgagataccacca 540
    || | | | | | | | | | | | | | | | | | | | | | | | |
DB 96269 CTGCATCTTACATTCGTGACGCTGCCAATGTGATGAGCACTTCTCTGTGATGATGCCCA 96210

```

QY	541	ttgttgctcttctgtctgttagccctgtgaatcaatgaagtgatgagtgatgttgcgcgat	600
Db	96209	CTGTAAAGCTCTTTGTGCTCCAAACCAATGTGACAGAGTTAAGGCTTGTGTCAGAT	96150
QY	601	attaccctggcaatgaagggaactttatcttctaccgcaatctcctaagtgttataatgtt	660
Db	96149	GTTGTTTCTGGCTGTAAGGAACTTCTCTGTATCATCCCTTCCTTAAGGTTTATTTGTGACC	96090
QY	661	gctattctcgtatccgcacagtagaagaaggaaggaagcctctcaaatgtctaatct	720
Db	96089	AGTATCCGAGATTAAGGTCATCAAGTAGAGGCAAGCAACGAGCTTTTCTTACCTGTTCTTCA	96030
QY	721	catctcacagatggtgaacccttactatctctctgtaaactaaactataccgcctgct	780
Db	96029	CATCTCTTTGTGTCACCATGTACTACATCTATCATCTTACCTTAATTTGGCCTTCA	95970
QY	781	tccagctcttaccttgaagaagacaagatggtgagctgtacactctatctcttgtgaatcc	840
Db	95969	TCCACTCTCTCTTAATACAGACAGCAAGGTGGTGTCCATCATCTTACACTTGGGTGACACC	95910
QY	841	acaattaaaccgatggtgtacagcttccaaatagaggagatgcaggcaggaattggna	899
Db	95909	ACCTTGAACCCCTCATCTACACTGTGAGAGCAAAAGATGTCAAGATGTGCACCTCCGAA	95851

RESULT 15	204143 bp	DNA	linear	HTG 20-DEC-2001
AC0096461				
LOCUS	Rattus norvegicus clone CH230-58J17, ***	SEQUENCING IN PROGRESS		
DEFINITION	***, 62 unordered pieces.			
ACCESSION	AC0096461			
VERSION	AC0096461.2	GI:17955456		
KEYWORDS	HTG; HTGS_PHASE1.			
SOURCE	Norway rat.			
ORGANISM	Rattus norvegicus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
REFERENCE				
AUTHORS	1 (bases 1 to 204143)			
	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alstirooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J., Benson,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,T., Bowles,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Butlay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dethorne,S.R., David,R., Davila,M.L., Davis,C., Davy,C., Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gablis,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsli,F., Howard,S., Huber,J., Huijck,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivel,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,B., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Nguyen,N., Nickerson,E., Nwokweno,S., Nguyen,A., Nguyen,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Ogih,M., Okunowo,G., Oragunye,N., Pickens,R., Primm,E., Pu,L.L., Peery,J., Perez,L., Peters,L., Pickens,R., Primm,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojupokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoshchik,N., Sisson,I., Sodergren,E., Sonikhe,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Swatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.			

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vetrà, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 204143)
Worley, K. C.

Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced gi:115628089.

Genome Center

```

Direct Submissions
Submitted (17-Sep-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15628089.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GPXC
Center clone name: CH230-58J17
----- Summary Statistics -----
Assembly program: Phrap, version 0.990329First call to
findPhrapList
Consensus quality: 156802 bases at least Q40
Consensus quality: 170292 bases at least Q30
Consensus quality: 181161 bases at least Q20
Estimated insert size: 177160; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; aarose-ip estimation
Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank-draft_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
9619: contig of 9619 bp in length
*
* 9620 9719: gap of unknown length
*
* 9720 18930: contig of 9211 bp in length
*
* 18931 19030: gap of unknown length
*
* 19031 27646: contig of 8616 bp in length
*
* 27647 27746: gap of unknown length
*
* 27747 33370: contig of 5624 bp in length
*
* 33371 33470: gap of unknown length
*
* 33471 41116: contig of 7646 bp in length
*
* 41117 41216: gap of unknown length
*
* 41217 48276: contig of 7060 bp in length
*
* 48277 48376: gap of unknown length
*
* 48377 54366: contig of 6060 bp in length
*
* 54367 54537: gap of unknown length
*
* 54537 61967: contig of 7430 bp in length
*
* 61967 62066: gap of unknown length
*
* 62066 65405: gap of 4339 bp in length
*
* 65405 66067: gap of unknown length
*
* 66067 73151: contig of 6646 bp in length
*
* 73151 73252: gap of unknown length
*
* 73252 79129: contig of 5878 bp in length
*
* 79129 79229: gap of unknown length
*
* 79229 85275: contig of 6046 bp in length
*
* 85275 85375: gap of unknown length
*
* 85375 85469: contig of 5034 bp in length
*
* 85469 90569: gap of unknown length
*
* 90569 94546: contig of 3977 bp in length
*
* 94546 94647: gap of unknown length
*
* 94647 99437: contig of 4791 bp in length
*
* 99437 99537: gap of unknown length
*
* 99537

```

* 99538 102811: contig of 3274 bp in length
* 102812 102911: gap of unknown length
* 102912 106305: contig of 3394 bp in length
* 106306 106405: gap of unknown length
* 106406 111977: contig of 5572 bp in length
* 111978 112077: gap of unknown length
* 112078 116021: contig of 3944 bp in length
* 116022 116121: gap of unknown length
* 116122 118407: contig of 2286 bp in length
* 118408 118507: gap of unknown length
* 118508 121535: contig of 3028 bp in length
* 121536 121635: gap of unknown length
* 121636 125078: contig of 3442 bp in length
* 125078 125177: gap of unknown length
* 125178 129179: contig of 4002 bp in length
* 129180 129279: gap of unknown length
* 129280 132170: contig of 2891 bp in length
* 132171 132270: gap of unknown length
* 132271 134079: contig of 1809 bp in length
* 134080 134179: gap of unknown length
* 134180 137358: contig of 3179 bp in length
* 137359 137458: gap of unknown length
* 137459 139978: contig of 2520 bp in length
* 139979 140078: gap of unknown length
* 140079 142972: contig of 2894 bp in length
* 142973 143072: gap of unknown length
* 143073 145874: contig of 2802 bp in length
* 145875 145974: gap of unknown length
* 145975 147768: contig of 1794 bp in length
* 147769 147868: gap of unknown length
* 147869 149593: contig of 1725 bp in length
* 149594 149693: gap of unknown length
* 149694 151287: contig of 1594 bp in length
* 151288 151387: gap of unknown length
* 151388 153421: contig of 1934 bp in length
* 153422 153422: gap of unknown length
* 153422 154471: contig of 2050 bp in length
* 154472 155571: gap of unknown length
* 155572 157991: contig of 2420 bp in length
* 157992 158091: gap of unknown length
* 158092 160551: contig of 2460 bp in length
* 160552 162741: contig of 2090 bp in length
* 162742 162841: gap of unknown length
* 162842 164428: contig of 1587 bp in length
* 164429 164528: gap of unknown length
* 164529 164771: contig of 1943 bp in length
* 164772 166571: gap of unknown length
* 166572 168429: contig of 1858 bp in length
* 168430 168529: gap of unknown length
* 168530 170496: contig of 1967 bp in length
* 170497 170596: gap of unknown length
* 170597 172644: contig of 2048 bp in length
* 172645 172744: gap of unknown length
* 172745 173655: contig of 1621 bp in length
* 173656 174465: gap of unknown length
* 174466 175806: contig of 1341 bp in length
* 175807 175906: gap of unknown length
* 175907 178082: contig of 2176 bp in length
* 178083 178182: gap of unknown length
* 178183 180094: contig of 1912 bp in length
* 180095 180194: gap of unknown length
* 180195 181465: contig of 1271 bp in length
* 181466 181565: gap of unknown length
* 181566 183890: contig of 2325 bp in length
* 183891 183990: gap of unknown length
* 183991 185397: contig of 1407 bp in length
* 185398 185497: gap of unknown length
* 185498 187581: contig of 2084 bp in length
* 187582 187681: gap of unknown length
* 187682 188760: contig of 1079 bp in length
* 188761 188860: gap of unknown length
* 188861 190644: contig of 1784 bp in length

* 190645 190744: gap of unknown length
* 190745 191885: contig of 1141 bp in length
* 191886 191985: gap of unknown length
* 191986 193129: contig of 1144 bp in length
* 193130 193229: gap of unknown length
* 193230 194840: contig of 1611 bp in length
* 194841 194940: gap of unknown length
* 194941 196464: contig of 1523 bp in length
* 196464 196563: gap of unknown length

Query Match 35.8%; Score 331.2; DB 2; Length 204143;
Best Local Similarity 60.0%; Pred. No. 3.9e-83;
Matches 552; Conservative 0; Mismatches 368; Indels 0; Gaps 0;

QY 4 aatcacagcgttgtaactagttacattatctcgtgacctacaaaagctgaactcag 63
DB 58436 AACCAAGACACGTGTAAACAGATGCTGCTTCAAGGCTTCTGAGATCGGACGATACAG 58495

QY 64 ggaattatctctcttcttctctatctgtctatctgtgcttctctctgcaaatgctc 123
DB 58496 CTGCTCTCTTTATCTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 58555

QY 124 atcatcattgccaataatcatagcaaaccttgcatagcccatgfatgttctctctg 183
DB 58556 ATCATCTTGCCATCATCCATGCAACCCCTGGGCTCCACACCCCATGATTTTCTTGTTC 58615

QY 184 acactgctgtgttgatcatcatctgcaacaagaatcatcatcgaagaatgctggacc 243
DB 58616 AATTGGGCTACCATGATGATCTCTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 58675

QY 244 atgctaacaagaanaatcatcatcatcatgagctgcatgtccagctctctgttc 303
DB 58676 CTGCTGTCAAAAGAAACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 58735

QY 304 acatgctctctgagctgagatggttctctctacacacatgacctatgacgctatg 363
DB 58736 ACATGCTGTGCTCTCTCAAGATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 58795

QY 364 gccattgttccctctctacatgactatcatatgaaacacacatgctgctgctg 423
DB 58796 GGCATCTGCACCTCTCTGATTCATGACGACCATGATGAGAAAGCATTTTGGCGGTG 58855

QY 424 ctcaagatgcatgcatgcatcatcatcatcatctctctgctgctgcatgcatgcat 483
DB 58856 GCTGCTGAGATGAGGCGGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 58915

QY 484 aggttgacttctgtggtgccaacacacacacacacacacacacacacacacacacac 543
DB -58916 CCTTGAATTTCTGTGAGTCCCAATGATTCATGACGACTTCTCTGAGAGTGCCTCC 58975

QY 544 ctggtcttctctctgagccctgtaagaatcaatgaggtgagtgatgctgcatatc 603
DB 58976 CTGCTTCTCTCTGTAGGCTCCACCTATGTAAGACATGTCATATGTTCTTGGCTCAG 59035

QY 604 accccgcatagggagcttattcttaccctgcatctcctcctgcttctctgct 663
DB 59036 TTTTATNGGATTTTGAATCTCTCTGATGACCATCGTGTATGCTTCTATCTCTCCAG 59095

QY 664 attctcgbatcgcagacagtagaagcaagagagccttctcaacatgcatcatcat 723
DB 59096 ATCTGGAAGATCGGACCTTCAAGAGGAGAAAGACCTTCTCAACGCTCTTCCAG 59155

QY 724 ctcaagtggtgagcccttactatctctctgtaaatcaactatatacgcctctctc 783
DB 59156 CTCACTGCTGTGTGATGATTAACACTGCTGCTTTTATGCTCAATATAAGCCCTGTCTCC 59215

QY 784 agctatacattgaaagagacaaggtgtagctgcatcattcctctgctgctcaca 843
DB 59216 AGCTATTAACGACGAGAAGCAAAATGCTGCTGTCTGTATACCATGCTGAGCCACAG 59275

QY 844 ttaaacccgagtggtgtaagcttccagaatagagagatgagcgaggaattagaaggtg 903
DB 59276 CTCAACCCCTGTGATCTACACTTGTGAGAAACAAGAGAGTCAAGGACGCTCTCAGAAATT 59335

Oy 904 ttgcatttctgaacacta 923
|| | ||||| | | | |
Db 59336 TTCCTTTCTCAGAAATTA 59355

Search completed: June 21, 2002, 12:36:47
Job time: 8112 sec

PA (SENO-) SENOMYX INC

XX Zozulya S;
 FI WPI: 2001-570867/64.
 DR P-PSDB: AAU24630.

XX Nucleic acids encoding human olfactory G protein-coupled receptors,
 PT useful for screening for compounds involved in olfactory sensation,
 PT where the compounds can be used in the food, pharmaceutical and
 PT cosmetic industries to customise odours -
 PS Claim 1; Page 135; 319pp: English.

XX The invention relates to nucleic acids encoding human olfactory
 CC receptors, OR, (a G protein-coupled receptor, GPCR). The ORs
 CC specifically recognise molecules, odorants, that elicit specific
 CC olfactory sensation. The human olfactory receptors and polynucleotides
 CC encoding them are useful for screening a library of chemical compounds
 CC for compounds that are involved in olfactory sensation. Modulators of
 CC their activity are useful for pharmacological and genetic modulation of
 CC olfactory signalling pathways. Therefore, they can be used in the food,
 CC pharmaceutical and cosmetic industries to customise odours and
 CC fragrances. The present sequence encodes a human olfactory receptor of
 CC the invention.

XX Sequence 924 BP; 219 A; 235 C; 184 G; 286 T; 0 other;

Query Match 99.7%; Score 920.8; DB 22; Length 924;
 Best Local Similarity 99.8%; Pred. No. 3.4e-276;
 Matches 922; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgaatcacagcgtgtgaactgaagtcattatctctggccctcccaaaaagcctgaac 60
 Db 1 atgaatcacagcgtgtgaactgaagtcattatctctggccctcccaaaaagcctgaac 60
 QY 61 cagggaattatct 120
 Db 61 cagggaattatct 120
 QY 121 ctcatcatcatgtgccaatctatagcaacacctgtgcatagcccatgtatgtttctt 180
 Db 121 ctcatcatcatgtgccaatctatagcaacacctgtgcatagcccatgtatgtttctt 180
 QY 181 ctgacacatggtctgtgtgacatcatctgcgaacacatcatcacaagatgctg 240
 Db 181 ctgacacatggtctgtgtgacatcatctgcgaacacatcatcacaagatgctg 240
 QY 241 accatgctacatcagaataacattcatatgcaaggctgcatgtccagctctcttg 300
 Db 241 accatgctacatcagaataacattcatatgcaaggctgcatgtccagctctcttg 300
 QY 301 ttacatgtctctggagctgagatggtctcttcacacacatgacctatgacgctat 360
 Db 301 ttacatgtctctggagctgagatggtctcttcacacacatgacctatgacgctat 360
 QY 361 gtggacattgttccctctctcatatgaactatgaacacacatgtgtgtagcc 420
 Db 361 gtggacattgttccctctctcatatgaactatgaacacacatgtgtgtagcc 420
 QY 421 ttgtcagcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatg 480
 Db 421 ttgtcagcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatg 480
 QY 481 atgaggtgactctctgtggccaacacacatgacacattctctctgtgagatacccca 540
 Db 481 atgaggtgactctctgtggccaacacacatgacacattctctctgtgagatacccca 540
 QY 541 ttgtcagcttgcctctgtagccctgtaagaataaggtgagtgatgtgtgtagt 600
 Db 541 ttgtcagcttgcctctgtagccctgtaagaataaggtgagtgatgtgtgtagt 600
 QY 601 attaccctggcagcatagggaacttattcttactgcatctcctatggtttatcaatg 660

Db 601 attaccctggcagcatagggaacttattcttactcgcgcatctcctatggtttatcaatg 660
 QY 661 gctattctccgtatccagcagtagaaggaaggaagggccttcacatgctcatc 720
 Db 661 gctattctccgtatccagcagtagaaggaaggaagggccttcacatgctcatc 720
 QY 721 catctcagctgtgacaccttacttactctcctgtatcatctacacattatccgctg 780
 Db 721 catctcagctgtgacaccttacttactctcctgtatcatctacacattatccgctg 780
 QY 781 tccagctatcatcttgaagaagacaaggtgtgactgcatccttactctgtgactcc 840
 Db 781 tccagctatcatcttgaagaagacaaggtgtgactgcatccttactctgtgactcc 840
 QY 841 acattaaacccgatgtgtacagcttccagataggaatgagcaggaattagaag 900
 Db 841 acattaaacccgatgtgtacagcttccagataggaatgagcaggaattagaag 900
 QY 901 gtgttgcattcttgaacacatag 924
 Db 901 gtgttgcattcttgaacacatag 924

RESULT 2
 AAH32037
 ID AAH32037 standard; DNA: 921 BP.

XX AAH32037:
 AC 30-JUL-2001 (first entry)
 DT
 XX

DE Human olfactory receptor polynucleotide, SEQ ID NO: 610.
 XX Human; olfactory receptor. OR: primary scent determination;
 KW secondary scent determination; polypeptide library; odour receptor;
 XX scent profile; scent fingerprint; scent representation; ds.

OS Homo sapiens.
 XX WO200127158-A2.
 PN 19-APR-2001.
 XX

PF 06-OCT-2000; 2000WO-US27582.
 XX
 PR 08-OCT-1999; 99US-0158615.
 PR 24-FEB-2000; 2000US-0184809.

XX (DIGI-) DIGISCENTS.
 PA (YEDA) YEDA RES & DEV CO LTD.
 XX

PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
 XX WPI: 2001-290713/30.
 DR

XX New polynucleotides which encode polypeptides involved in olfactory
 PT sensation for identifying olfactory agonists and antagonists -
 XX
 PS Claim 8; Page 425; 1857pp: English.

XX The present sequence is one of a number of isolated polynucleotides
 CC which encode polypeptides involved in olfactory sensation. The
 CC polynucleotides can be used in screening for olfactory agonists and
 CC antagonists. The methods allow for the determination of primary
 CC scents and the identification of the odour receptors used to detect
 CC these primary scents. The methods also enable determination of
 CC secondary scents and the identification of combinations of odour
 CC receptors that are involved in detecting such secondary scents.
 CC This enables the construction of a scent representation (also called
 CC a scent fingerprint or scent profile), which may be used to re-create
 CC and edit scents. Libraries of olfactory receptors are useful for
 CC determining the interaction pattern of a composition with the receptors,

CC and antibodies raised against them. The NOV1-11 polypeptides and nucleic
CC acids are used to identify ligands and ant/agonists, they are also used
CC to diagnose and treat GPCR-associated diseases, e.g. cardiomyopathy,
CC atherosclerosis, diseases related to cell-signalling or metabolic
CC pathways, bacterial, viral, fungal or protozoal infections, pain,
CC anorexia, bulimia, asthma, neurological disorders (e.g. Parkinson's and
CC Alzheimer's diseases) stroke, multiple sclerosis, seasonal
CC psoriatic skin, ischaemia, chirotnic hepatitis, acute pancreatitis,
CC diabetes, cancer, angioneuromas, obesity, also disorders of olfactory
CC function. The nucleic acids can be used as marker for chromosome 11,
CC therapeutically, including as antisense or ribozyme sequences, to prepare
CC transgenic animals, as sources of primers and probes (e.g. for
CC diagnosis), for recombinant production of the proteins, including in gene
CC therapy, and for tissue typing. The present sequence encodes human NOV6.
XX Sequence 798 BP; 191 A; 202 C; 164 G; 241 T; 0 other;
SQ

Sequence 798 BP; 191 A; 202 C; 164 G; 241 T; 0 other

Query Match	84.3%	Score 778.8, DB 24;	Length 798;
Best Local Similarity	99.7%	Pred. No. 5.2e-232;	
Matches 780; Conservative	0; Mismatches	2; Indels	0; Gaps

QY	143	atagcaaacacctgtgatagcccatbtagtttctctctgtgcaactggtgttgaggaca	202
Db	1	atacaacaacactgtgcatacggcccaatgtaagtttctctctgtgcaactggtgttgaggaca	60
QY	203	tcatctgcacaacaagaagcatatacgcagaagatgtctggaggacatgctaaataagaanaata	262
Db	61	tcatctgcacaacaagaagcatatacgcagaagatgtctggaggacatgctaaatacagaanaata	120
QY	263	ccatttcaatgtgaaggtgtgatgtcccaagctctctctgttcaaatggtctctvgagactg	322
Db	121	ccatttcaatgtgcagctgtcatgtgtccagactctctgttcaaatggtctctctvgagactg	180
QY	323	agaagtgctctcttccacaacacatgagctctatgacgctatgtggccatttgtttccctctc	382
Db	181	agaagtgctctcttccacaacacatgagctctctatgacgctctatgtggccatttgtttccctctc	240
QY	383	attcaagracattatgataaacacacatgtgtgtgagctgtgtcagaatgagatgacta	442
Db	241	attcaagtaacggttatgataaacacacatgtgtgtgagctgtgtcagaatgagatgacta	300
QY	443	ttagcagtcaccaatctcctgggtgtgcacacagctcttatacagaaggtttacttctgtgtgc	502
Db	301	ttagcagtcaccaatctcctgggtgtgcacacagctcttatacagaaggtttacttctgtgtgc	360
QY	503	caaaacacatgtgacacattctctctgtgtgatacccaatgtgcgcttgttctgtatgcc	562
Db	361	caaaacacatgtgacacattctctctgtgtgatacccaatgtgcgcttgttctgtatgcc	420
QY	563	ctgtgaagaatcaatgaggtgtgatgtgtatgttgcgtatatacccttggccataagaggact	622
Db	421	ctgtgaagaatcaatgaggtgtgatgtgtatgttgcgtatatacccttggccataagaggact	480
QY	623	ttaatttcaacgcgcatctctctatctgtgtttatataatgtgtgtacttctccgatatccgacag	682
Db	481	ttaatttcaacgcgcatctctctatctgtgtttatataatgtgtgtacttctccgatatccgacag	540
QY	683	tagaagagcaagaggaagcctctccaacttggttaactatatactatccacagtggtgacacctt	742
Db	541	tagaagagcaagaggaagcctctctccaacttggttaactatatactatccacagtggtgacacctt	600
QY	743	acatattctcgttaactatacctataatccgcgcctgtctccagctatacatattgaaagag	802
Db	601	acatattctcgttaactatacctataatccgcgcctgtctccagctatacatattgaaagag	660
QY	803	acaagtggtgtagtgcaactatactctgtgtgactccacatataaacccagatggttataa	862
Db	661	acaagtggtgtagtgcaactatactctgtgtgactccacatataaacccagatggttataa	720
QY	863	gcttcacgaataagagagatgcagtcagaagaattaggaagtggttgcattctvgaaacct	922
Db	721	gcttcacgaataagagagatgcagtcagaagaattaggaagtggttgcattctvgaaacct	780

QY	923	ag	924
Db	781	ag	782

RESULT	4
AAS42284	
ID	AAS42284 standard; cDNA; 987 BP

AC AAS42284

DT 18-DEC-2001 (first entry)

XX
DE Human cDNA encoding olfactory receptor AOLF81.

XX Human; olfactory receptor; G protein-coupled receptor; GPCR; odourant
KW food additive; cosmetic; fragrance; pharmaceutical additive.

OS Homo sapiens

PN WO200168805-A2

PD 20-SEP-2001.

13-MAR-2001; 2001WO-US07771.

AA
PR 13-MAR-2000; 2000US-0188914

PR 12-APR-2000; 2000US-0198474

PR 26-MAY-2000; 2000US-0207702

PR 23-JUN-2000; 2000US-0226534
PR 16-AUG-2000; 2000US-0226534

07-FEB-2001; 2001US-0266862

XX
PA (SENO-) SENOMYX INC.

XX
PI ZOZULYA S;

XX
DB WPT: 2001-570867/64.

DR P-PSDB; ANOZETA;
XY

PT Nucleic acids encoding
PT useful for screening

PT where the
cosmetic

XX 0134w 1: Page 117: 319pp: English.

XX
XX
The invention re

CC receptors, OR, (a G protein-coupled receptor, that elicit specific

olfactory sensation. The human olfactory system is useful for screening

CC for compounds that are involved in olfactory sensation; CC for compounds that are useful for pharmacological and genetic modulation of olfactory sensation.

olfactory signalling pathways. Therefore, they can be used in a wide range of industries to customise products and services.

xx Sequence 987 BP; 190 A; 309 C; 235 G; 253 T; 0 other;
SQ

QY 4 aatcacgcgttgaactgagtcattatcttgycctcccaaaaagcctgaaccacg 63
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67 aaccagacgttgyaacgcagttcatctcagggctttcgagacaccagaataacgg 126

[illegible]

KM	viridic: antiinflammatory; anticonvulsant; antileptic; neuroprotective; neurotrophic; cerebroprotective; hypotensive; tranquilizer; vulnery;
KM	ophthalmological; cell proliferative disorder; actinic keratosis;
KM	anorectic; arteriosclerosis; atherosclerosis; cirrhosis; hepatitis;
KM	psoriasis; cancer; neurological disorder; stroke; Alzheimer's disease;
KM	Huntington's disease; Parkinson's disease; cardiovascular disorder;
KM	epilepsy; hypertension; varicose vein; vasculitis; dysphagia; dyspepsia;
KM	anorexia; gastrointestinal disorder; pancreatitis; autoimmune disorder;
KM	Addison's disease; Crohn's disease; acquired immune deficiency syndrome;
KM	AIDS; uveitis; infection; trauma; metabolic disorder; diabetes; obesity;
KM	osteoporosis; transgenic animal; gene therapy; ss.
OS	Homo sapiens.
PH	Key
FT	Location/Qualifiers
FT	253..1241
FT	CDS
FT	253..1241
FT	/tag= a
FT	/product= "Human GCRC-9 protein"
PN	W0200157085-A2.
PD	
XX	09-AUG-2001.
XX	
PF	01-FEB-2001; 2001WO-US03455.
XX	
PR	02-FEB-2000; 2000US-0180093.
PR	11-FEB-2000; 2000US-0182045.
XX	
PA	(INCY-) INCYTE GENOMICS INC.
PI	
XX	Baughn MR, Au-Young J, Yue H;
XX	
DR	WPI: 2001-488869/53.
XX	P-PSDB: AAE06759.
PT	
PT	Novel isolated human G-protein coupled receptor useful for diagnosing,
PT	preventing and treating cell proliferative, neurological,
PT	cardiovascular, gastrointestinal, autoimmune/inflammatory and metabolic
PT	disorders -
XX	
PS	Claim 5; Page 133; 138pp; English.
XX	
CC	The present sequence is human G-protein coupled receptor-9 (GCRC-9)
CC	cDNA. The present invention relates to GCRC protein and nucleic acids
CC	encoding them. GCRC protein, its agonist or antagonist are useful for
CC	treating diseases or conditions associated with decreased expression
CC	or overexpression of functional GCRC in a patient, where the disorder
CC	is selected from cell proliferative disorders such as actinic keratosis,
CC	arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, psoriasis, and
CC	cancer, neurological disorders such as epilepsy, stroke, Alzheimer's
CC	disease, Huntington's disease, Parkinson's disease, cardiovascular
CC	disorders such as hypertension, vasculitis, varicose veins, gastro-
CC	intestinal disorders such as dysphagia, dyspepsia, anorexia, nausea,
CC	pancreatitis, autoimmune/inflammatory disorders such as acquired
CC	immunodeficiency syndrome (AIDS), Addison's disease, Crohn's disease,
CC	uveitis, viral, bacterial, fungal, parasitic, protozoal, helminthic
CC	infections, trauma and metabolic disorders such as diabetes, obesity,
CC	osteoporosis. GCRC proteins and their cDNAs are used to assess the
CC	effects of exogenous compounds on the expression of GCRC sequences.
CC	GCRC cDNA is useful to create knock in humanised animals (pigs) or
CC	transgenic animals (mice or rats) to model human disease, for
CC	therapeutic or diagnostic purposes, for somatic or germline gene
CC	therapy, to generate hybridisation probes useful in mapping the
CC	naturally occurring genomic sequence, and in molecular biological
CC	techniques.
XX	
90	Sequence 1241 BP; 259 A; 366 C; 295 G; 321 T; 0 other;

Query Match	33.2%;	Score 307.2;	DB 22;	Length 1241;
Best Local Similarity	58.4%;	Pred. No. 4.9e-85;		
Matches 537;	Conservative 0;	Mismatches 383;	Indels 0;	Gaps 0;

QY 4 aatcacagctgttaactgaattcatattctctgggctccaccaaagcctgaactccag 63
DB 321 aaccagagctgttaacagagctcctcctgagagcttcttgagagcccccgaattaccgg 380
QY 64 ggaattatct 123
DB 381 ggtctcttattgaactgttctctctctctctctctctctctctctctctctctctct 440
QY 124 atcatcattggcaaaatcatatagcaacacctgtgtacagccatgtatgtttctctctg 183
DB 441 atcaactcttgccatcaactgtcaacacctgtgtgtccacagctcctattgttcttctac 500
QY 184 acactgt 243
DB 501 aactgt 560
QY 244 atgtcaacatcagaataatcatattcatatgtcagagctgtatgtccagctctctctgttc 303
DB 561 ctgt 620
QY 304 acatgt 363
DB 621 acgt 680
QY 364 gcaattgt 423
DB 681 gcaatctgt 740
QY 424 ctacagatgt 483
DB 741 gccacacgcgt 800
QY 484 aggt 543
DB 801 cgctgt 860
QY 544 ctgt 603
DB 861 ctgt 920
QY 604 accctgt 663
DB 921 ttctacagatgt 980
QY 664 attctcgtatccgacagtagaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 723
DB 981 attcctgaaggtgaagactgtcctgt 1040
QY 724 ctcaacagctgtgaaccttactatctctctgtatctcaacctatatacctatatacctgtgtgtgt 783
DB 1041 ctcaacagctgt 1100
QY 784 agctatcatattgaagagagaaggt 843
DB 1101 ggtctacagcgaaggaaggaaggaaggt 1160
QY 844 ttaaacccagctgt 903
DB 1161 ctcaacccctcatatattgtgagaacaaggaaggtcaaaagcagccctcaggaagctt 1220
QY 904 ttgcattctgaacacta 923
DB 1221 ttcccttctcagaactta 1240

RESULT 6
AAH31706
ID AAH31706 standard; DNA; 924 BP.

XX AAH31706;
XX AC
XX DN 30-JUL-2001 (first entry)
XX

DE Human olfactory receptor polynucleotide, SEQ ID NO: 279.
XX Human: olfactory receptor; OR: primary scent determination;
KW Secondary scent determination; polypeptide library; odour receptor;
KW scent profile; scent fingerprint; scent representation; ds.
XX Homo sapiens.
XX WO200127158-A2.
XX 19-APR-2001.
XX 06-OCT-2000; 2000MO-US27582.
XX 08-OCT-1999; 99US-0158615.
XX 24-FEB-2000; 2000US-0184809.
XX (DIGI-) DIGISCENTS.
XX (YEDA) YEDA RES & DEV CO LTD.
XX Bellenson J, Smith D, Lancel D, Glusman G, Fuchs T, Yanai I;
XX WPI: 2001-290713/30.
XX New polynucleotides which encode polypeptides involved in olfactory
XX sensation for identifying olfactory agonists and antagonists -
XX Claim 8; Page 289; 1857pp; English.
XX The present sequence is one of a number of isolated polynucleotides
XX which encode polypeptides involved in olfactory sensation. The
XX polynucleotides can be used in screening for olfactory agonists and
XX antagonists. The methods allow for the determination of primary
XX scents and the identification of the odour receptors used to detect
XX these primary scents. The methods also enable determination of
XX secondary scents and the identification of combinations of odour
XX receptors that are involved in detecting such secondary scents.
XX This enables the construction of a scent representation (also called
XX a scent fingerprint or scent profile), which may be used to re-create
XX and edit scents. Libraries of olfactory receptors are useful for
XX determining the interaction pattern of a composition with the receptors,
XX and can be used for determining differences in the olfactory faculties
XX of different individuals.
XX Sequence 924 BP; 170 A; 293 C; 219 G; 242 T; 0 other;
SO
Query Match 33.1%; Score 305.4; DB 22; Length 924;
Best Local Similarity 58.4%; Pred No. 1.5e-84;
Matches 534; Conservative 0; Mismatches 381; Indels 0; Gaps 0;
QY 4 aatcacagctgttaactgaattcatattctctgggctccaccaaagcctgaactccag 63
DB 10 aaccagagctgttaacagagctcctcctgagagcttcttgagagcccccgaattaccgg 69
QY 64 ggaattatct 123
DB 70 ggtctcttattgaactgttct 129
QY 124 atcatcattggcaaaatcatatagcaacacctgtgtacagccatgtatgtttctctctg 183
DB 130 atcaactcttgccatcaactgtcaacacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 189
QY 184 acactgt 243
DB 190 aactgt 249
QY 244 atgtcaacatcagaataatcatattcatatgtcaggtgtgtgtgtgtgtgtgtgtgtgtgtgt 303
DB 250 ctgt 309
QY 304 acatgt 363

QY 661 gctattctccgataccgacagtagaaggaaggaagccttcaacatgctcatct 720
DB 673 gctgtgtgaagatcagctcagctacaggtacgaagcctttagactgtgacct 732
QY 721 catctacagtgtagccttcttacttctcctgttaactatatacctatccgctct 780
DB 733 cactgtgactgctgtgacctcttctatggttctgtatctcatgtacatgcgacct 792
QY 781 tccagctatacatgtgaagaaagagtgagctgacgtctatctactgttgacctcc 840
DB 793 tccagctactccttaacacaggaagagtggtccatctctatgacctgtgacctcc 852
QY 841 acattaaaccgagtgtagcagcttccagaaataggagatgcaaggagaaataggaa 900
DB 853 gtgtgtgaatcccatctcatcactttagaataagagataaataatgcataagaa 912
QY 901 g 901
913 g 913
RESULT 8
AAS42360 8
ID AAS42360 standard; cDNA, 975 BP.
XX AAS42360:
AC AAS42360:
XX 18-DEC-2001 (first entry)
DT Human cDNA encoding olfactory receptor AOLFRI65.
DE Human cDNA encoding olfactory receptor; G protein-coupled receptor; GPCR; odourant;
XX Human; olfactory receptor; G protein-coupled receptor; GPCR; odourant;
KW bs; food additive; cosmetic; fragrance; pharmaceutical additive.
KM Homo sapiens.
XX MO200168805-A2.
PN 20-SEP-2001.
PD 13-MAR-2001; 2001WO-US07771.
PE 13-MAR-2000; 2000US-0188914.
XX 13-MAR-2000; 2000US-0188914.
PR 12-APR-2000; 2000US-0192033.
XX 12-APR-2000; 2000US-0192033.
PR 24-APR-2000; 2000US-0193335.
XX 24-APR-2000; 2000US-0193335.
PR 26-MAY-2000; 2000US-0207702.
XX 26-MAY-2000; 2000US-0207702.
PR 23-JUN-2000; 2000US-0213849.
XX 23-JUN-2000; 2000US-0213849.
PR 16-AUG-2000; 2000US-0226534.
XX 16-AUG-2000; 2000US-0226534.
PR 07-SEP-2000; 2000US-0230732.
XX 07-SEP-2000; 2000US-0230732.
PR 07-FEB-2001; 2001US-0268662.
XX (SENO-) SENOMYX INC.
PA Zozulya S:
XX WPI: 2001-570867/64.
DR P-PSDB: AAU24667.
XX Nucleic acids encoding human olfactory G protein-coupled receptors,
PT useful for screening for compounds involved in olfactory sensation,
PT where the compounds can be used in the food, pharmaceutical and
PT cosmetic industries to customise odours -
XX Claim 1; Page 151; 31pp; English.

CC olfactory signalling pathways. Therefore, they can be used in the food,
CC pharmaceutical and cosmetic industries to customise odours and
CC fragrances. The present sequence encodes a human olfactory receptor of
CC the invention.
XX
SQ Sequence 975 BP; 208 A; 252 C; 210 G; 305 T; 0 other;
Query Match 28.9%; Score 267.4; DB 22; Length 975;
Best Local Similarity 56.08; Pred. No. 1,1e-72;
Matches 505; Conservative 0; Mismatches 396; Indels 0; Gaps 0;
QY 1 atgatacaacgctgtgaactgagttcaattatctggcctccacaaagcctgactc 60
DB 13 aggaacaacaacattgtgacaaatattctctcctggactttagaccctcaaatg 72
QY 61 cagggaattatctccctttttctcatgtgcatctgttggtttcttcggcaatg 120
DB 73 aagatttccctttcaatgtattcttctgggctcactcctctgaagtggtcggaactta 132
QY 121 ctcatcatcttgccaaatctatagacaacacttgataagccatgtatgtttcct 180
DB 133 agcctcatctgcccattataagatgactctcactcctgacatgacatgacttctctc 192
QY 181 ctgacactggtgtgtgtgacatcatctgcacaacagcatcatcaggaatgtcg 240
DB 193 agtaacctgtcctctccttgatctgcatgtgltcctccaccgcccacaaatgtcgtc 252
QY 241 accatgcttaacatacgaagaataacattcatatagagagctgcatgtgccagctctctg 300
DB 253 gacatcatcagagagcagaagaacacattctcctgtgtgctgcccactgactgtgc 312
QY 301 ttcacatggtctctggtgagctgagatgtctcttcacacacatggtcctatgacgctat 360
DB 313 tctctggtgagatggtgctgagctgaatgcttctccggagctatgagcctgaacggtat 372
QY 361 gtggccattgtttcccttctcattacagtaactatatagaacacacatgtgtgtagcc 420
DB 373 gctgcaatctgaaacccctgttacaagaagcccatataccatacactgtgttaag 432
QY 421 ttgctagaatgtgcatgtgctatgtcagtcaccaatctcctggtgtcacagctctatc 480
DB 433 atgtgtgtgtggccctatgtgtgattcctttagtcttcaatgaaacataactctgtc 492
QY 481 atgaagttgaacttctgtgtgtgccaacaccattgacacttctctgtgatacccca 540
DB 493 tatcagcatgattctgtgtgtgcccataatgaatcaacacatttctgtgacctcca 552
QY 541 ttgtgtgcttgtcctgtgagccctgtaagaatcaatgagtgatgtgtatgtctgat 600
DB 553 gtccctgctctgtccctgtgactctgacccacagcaggtgtgtgaccttaagtcagt 612
QY 601 attacctgacataagggagacttattcttaactgcatctcatctcatgtgtttatgtt 660
DB 613 gttgtcgttggaatagtgctgtgtcagtggttccatcctctatagttatattgttgt 672
QY 661 gctattctccgataccgacagtagaaggaaggaagccttcaacatgctcatct 720
DB 673 gctgtgtgaagacatcctcagctacaggtaggaacaaaggcctttagcactgtgacct 732
QY 721 catctacagtgtagccttcttacttctcctgttaactatatacctatccgctctg 780
DB 733 cactgtgactgctgtgacctcttctatggttctgtacttctatgacagcagcaagt 792
QY 781 tccagctatacatgtgaagaaagagtgtagctgacgtctatctactgtgactcc 840
DB 793 tccagctactccttaacacaggaagagtggttccatctctatctatgactgtgacctcc 852
QY 841 acattaaaccgagtgtagcagcttccagaaataggagatgcaaggagaaataggaa 900
DB 853 gtgtgtgaatcccatctcatcactttagaataagagataaataatgcataagaa 912
QY 901 g 901

[illegible]

Query Match	28.23;	Score 260.8;	DB 22;	Length 930;
Best Local Similarity	55.78;	Pred. No. 1.2e-70;		
Matches 499;	Conservative	0;	Mismatches 397;	Indels 0;
Gaps 0;				
14	tttgaactgaattcattatcttctgggcctccacaaagcctgaactccaggaattatct	73		

Db	23	tggtagctgagttatcttcattgagcgtacacccagagaccctcagatggaggtacatctct	82
Oy	74	tcctctttttctccatgtgtctactctgttgcgtcttcttcggcaaatgctcatcatgt	133
Db	83	tcgtgtgtcttccctcagatttaacctgttaactgttagtgggagatatlgttatgatacc	142
Oy	134	ccaaatctatagcaaacacctgtgataagccgatgtatgtttctcttcgtgacactgctg	193
Db	143	tgtattaaacagacactcagcttccacacaccatgtattttctcttcgcgaacctctct	202
Oy	194	tttggagacatcatctgcacacaaagcatcatacggaagatgtcttggggagccatgtcaacat	253
Db	203	ttgtttgccctgggtactactctctcagccatgtgcccocaggagatgtgtgctgacttccaaaca	262
Oy	254	cagaaaataaccattcatatgacagctgcgatgtccagctctctctgttcaaatgtgtc	313
Db	263	atccacaagttatctctctctccagctgtgcccacccagttgcttttlttgaagtttg	322
Oy	314	tggtagcttgtagtggtctctcttccacacaaatggccctcatgacgcgtatgtgagccattgt	373
Db	323	tggatgtcgtgagtgctagctgcttcggcagccaatgagccatgctatgtctgttttgggscattgtc	382
Oy	374	tcctctcttcaataagctactatattgaacccacatatgttgttgaagctgtgcagatg	433
Db	383	gacccttcacatatagacacttcaatgcccagaaggtctgctgtctcactcagctgggct	442
Oy	434	tcatgtctattgacggtacccaattccctgggttgacacagctcttactcatgaggttgaact	493
Db	443	cttaccctgtgcgtctcagtgagtttagtaagcccacactacactcactcagccctgagtt	502
Oy	494	tctgtggcccaaacacatgtgaccactctctctgtgagabaccccatgtgctgtgt	553
Db	503	actgtgttccaatatatacatcatcatcttctcttctgcgaaatccacacactcttggccctct	562
Oy	554	ccgtgagccctgtagaatcaatgaggttgatgtgtatgttgcgtatcatccctggcca	613
Db	563	ctgtgtcagacacactcaacgtgagagcttgcctctctcctcgaatcgtgtgtctcatgtaat	622
Oy	614	taggagacttattcttcaaccctgcacctccctatggtttatcaatltgttcatctccgta	673
Db	623	tcagacacatcctcatcatcttcatctccatataacttcatcttgttgaatcatagaaa	682
Oy	674	tcgcgacagttagaagcgcaagaaggagcccttcaaaatgctcatcctcatcagacgttg	733
Db	683	tgcgttcacgtgtaaaggccgccttaaggctcttccaccctggggtctcaacttaactgca	742
Oy	734	tgaaccttactatctctctgttaactacaactataccgcgcctgtcttcagctatacat	793
Db	743	tcaacctctctcatgtgacacagtcatagtttatgtactcgtgaagccaacacacagttactccc	802
Oy	794	ttgaaagagacaaggtgtgtagctgcactctatactctgtgagccccaacataaaccca	853
Db	803	tggacacagacacagtggtgccccttggtcttcaacacagttatcaatcccacatgttaaacctc	862
Oy	854	tgggttataagcttccagatatggagagatgacgtgcagagcaggaattagaaagtgltttga	909
Db	863	tgattctacagtttgcggacaagaagtgtgaaagctgtcttcaaaaagctaattgga	918
RESULT 11			
AAAS1177			
ID	AAAS1177 standard; cDNA; 669 BP.		
XX	AAAS1177;		
XX	26-SEP-2000 (first entry)		
XX	Murine olfactory receptor ligand-binding region cDNA 3.		
KW	Degenerate primer: ligand-binding region; olfactory receptor;		
KW	Transmembrane domain; N-terminal; plasma membrane; translocation domain;		
KW	human rhodopsin receptor; odorant; toxicity; olfactory response; ss.		
XX			

DT	26-SEP-2000 (first entry)
XX	
DE	Murine olfactory receptor ligand-binding region cdna 3

Degenerate primer; ligand-binding region; olfactory receptor; Transmembrane domain; N-terminal; plasma membrane; translocation domain; human rhodopsin receptor; odorant; toxicity; olfactory response; ss.

OS	Mus musculus.
XX	
FH	Key
EH	Location/Qualifiers
FT	1..669
CDS	/*tag= a
FT	/product= Olfactory_receptor_ligand_binding_protein
FT	/partial
FN	
PX	WO200035274-A1.
XX	
PD	22-JUN-2000.
XX	
PF	17-DEC-1999; 99WO-US30221.
XX	
PR	17-DEC-1998; 98US-0112605.
XX	
PA	(UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
PI	
DR	Reed RR, Krautwurst D, Yau KW;
XX	
DR	WPI: 2000-431471/37.
XX	P-PsDB; AAY96664.
PT	
XX	primer pair for amplifying olfactory receptor nucleic acid, useful e.g.
XX	for producing receptor libraries used in e.g. screening odorants for
PS	toxicity
XX	
PS	Claim 18; Page 31; 61pp; English.
XX	
CC	AAY51175-93 encode murine odorant/ligand binding regions of olfactory
CC	receptors. They were amplified using degenerate primers shown in
CC	AAY51163-64.
CC	The primers can be used to generate a library of OR-LBR comprising the
CC	nucleic acids domains (TMD) II-VI, III-VI, IV-VII or III-VIII. Chimeric
CC	transmembrane domains (TMD) II-VI, III-VI, IV-VII or III-VIII. Chimeric
CC	nucleic acids encoding a 7-TMD protein comprise a sequence encoding an
CC	N-terminal plasma membrane translocation domain (especially the sequence
CC	shown in AAY9658), a first TMD and the OR-LBR. The translocation domain
CC	was initially derived from the N-terminus of the human rhodopsin
CC	receptor. LBR produced by amplification with the primers are useful for
CC	generating new odorants, to screen for toxicity or therapeutic activity
CC	in odorants, and altering an animal's olfactory response (claimed).
XX	
XX	
SQ	Sequence 669 BP; 131 A; 200 C; 142 G; 196 T; 0 other;
Query Match	28.1%; Score 259.8; DB 21; Length 669;
Best Local Similarity	62.0%; Pred. No. 2e-70;
Matches 411; Conservative	0; Mismatches 252; Indels 0; Gaps 0;
OY	187 ctgctgttcttgagacatcatctgcacaagaacatcaccgaagtctggggacatg 246
b	
7	tgcgcaccacatgatattatctcgacccttcctctgtgctgcgccaaaggcgtgttgtcta 66
OY	247 ctaaacataagaanaaatcacattttcaatatgcaggtgcattgcccactcttctgttaca 306
y	
Dd	67 ctatctggagaaaacacacatctccctttaaagggtgcatgycgcacgtctcttccctgtg 126
OY	307 tggctctcggagagcgagatggttctcttcaccacacatggccttatgacgcgtatgtgccc 366
Dd	127 tggctcctgtcttcacagacctgctgctgtctccaagtcattgagccgtcattgtggcc 186
OY	367 attgttccctctcatattacaglaactatattgaaccaacacatagtgtgtgacttggc 426
Dd	187 atctgcttccccctgactacacagctcctgaagtggccacagttgtgtgggctctggcc 246
OY	427 agcattgcatggcgcattgcagtcaccacattctctgtggcgacaacagctcttatcatgag 486
Dd	247 atcggatgatggtccatctgtgctctggaatgatcatatcaaacctggtctgtatgacacg 306
OY	487 ttgacttctctggggccaacacacattgacaccttctctgtgagataaccacattgtcg 546
Dd	307 ctgtcatctctgtggaccacaaaggatcataccacactctctctgtgagatctccacactctt 366

Oy	547	gcttctgctctgtagccctcgtaagaatcaatgaagtgatgtgtgtatcgtgtcgtatatacc	606b
Db	367	ctgcctctctgttagcccccacacacgtaaacagcatatgatcctaaatgaagaatgtcttc	426b
Oy	607	ctggcacaagaaggacatttatcttaccctgcacccaccatagtgtttatcatgtgtccatc	666b
Db	427	tatggaaagcatcaattttgtgtcttacccttactatcccatgtgcgtgcacatctggcagatc	466b
Oy	667	ctccgattccgcacagatgaagaagcaagaaggcccttcctcaaaatgcatacctcatctc	726b
Db	487	ctgcgcacatgcgtctctgcgtgaggaaggaaggccctttctaccctgtcatccacactc	546b
Oy	727	acagtggtgagacccttctactatctccctcgttaactcaaacctataccgcctcgtctcagc	786b
Db	547	atcgtggctctctgtgtatctactcatctgtgtcttcgtccctatgcaaccctcgtaactcagc	606b
Oy	787	tatacatcttgaagaagacaagtgtagctgcactctatactcttcttgagactccacata	846b
Db	667	tatagcccaagaagaagaagaagttaactctgtgtgtatcatctatccctcagcccaacctg	666b
Oy	847	aac	849
Db	667	aac	669
RESULT 12			
AAH31926			
ID	AAH31926 standard; DNA; 975 BP.		
XX	AAH31926;		
DT	30-JUL-2001 (first entry)		
DE	Human olfactory receptor polynucleotide, SEQ ID NO: 499.		
KM	Human: olfactory receptor; OR: primary scent determination;		
KW	secondary scent determination; polypeptide library; odour receptor;		
XX	scent profile; scent fingerprint; scent representation; ds.		
OS	Homo sapiens.		
PN	MO200127158-A2.		
PD	19-APR-2001.		
PF	06-OCT-2000; 2000MO-US27582.		
PR	08-OCT-1999; 99US-0158615.		
PR	24-FEB-2000; 2000US-0184809.		
XX	(DIGI-) DIGISCENTS.		
PA	(YEDA) YEDA RES & DEV CO LTD.		
XX	Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;		
DR	WPI; 2001-290713/30.		
XX	New polynucleotides which encode polypeptides involved in olfactory		
PT	sensation for identifying olfactory agonists and antagonists -		
XX	Claim 8; Page 380; 1857pp; English.		
XX	The present sequence is one of a number of isolated polynucleotides		
CC	which encode polypeptides involved in olfactory sensation. The		
CC	polynucleotides can be used in screening for olfactory agonists and		
CC	antagonists. The methods allow for the determination of primary		
CC	scents and the identification of the odour receptors used to detect		
CC	these primary scents. The methods also enable determination of		
CC	secondary scents and the identification of combinations of odour		
CC	receptors that are involved in detecting such secondary scents.		
CC	This enables the construction of a scent representation (also called		
CC	a scent fingerprint or scent profile), which may be used to re-create		
CC	and edit scents. Libraries of olfactory receptors are useful for		

Db 172 atcattgtgctatcagcttgatgataccttcatccccaatgatactctcttcgcc 231
 QY 184 acactgtgctgttggaatcatctgcacaaagcatataccgaagatgctggagcc 243
 Db 232 aatctatctcttgatgatacttccatcttccaaactcagctcccaaatgctgagaa 291
 QY 244 atgctacatcgaanaatccattcatatgcaagctgcatgcccagctctctgttc 303
 Db 292 attcaacaaagatcaatccatctcttatgagagctgatacaacagatgtactttct 351
 QY 304 acatgtctctgagagctgagatggtctctcccaacatgagctatgacagctatg 363
 Db 352 atgtgtctgtcgtcatctgacaattgctcttgaggagcatggtcatgaccacttg 411
 QY 364 gccattgttccctcttcatcagtaactatgatgaaccacatgtgtgtagcctg 423
 Db 412 gcatctgcaccccttgatatacatatcgaatgagccaggttgcattgtgc 471
 QY 424 ctcaagatgtaactgctatgcaagcacaattcctggtgagacagctctatcatg 483
 Db 472 acagtcatactatgcttctcagtaataatattgtctgcgaacacaccccttgccalt 531
 QY 484 aggttgactctctgtggccaacacacatgcaactctctctgtgagatacccccattg 543
 Db 532 caattgctctctgtgaacacaacactctcccaactctctctgtgacttgccctctg 591
 QY 544 ctggctgtctctgtgagccctgtaagaatgaatgagtgatgtgtatgtctgalt 603
 Db 592 ctcaaatctctctgtcagatcatatgcaatgagctgtgttcttattgttggttta 651
 QY 604 accctggcagtagggagacttattcttaccatcactcatgattatgattatgttgc 663
 Db 652 tcaatatacatcttcccttcaactcagctctcttcttctatgctgcatcagaagct 711
 QY 664 attctcgtatcgcacagtagaaggaggaagggcctctcacaatgctcatcact 723
 Db 712 gctcgtagagatcttccacacaggaagtgaaagcctctcactgtgctctcact 771
 QY 724 ctcaacagtggtgaaccttactatctctctgtgaatcaactatactccgcccctgc 783
 Db 772 ctgaacagtgatacttactgtcttcaagaaactgtgagcggttacttctccctctcc 831
 QY 784 agctatatactgaaagagacagatgtgagctgcaactcatatcctgttgcctcaca 843
 Db 832 actcaccctgagagacagcagatagatgtgtcttccatcactcagtggtgacacccatg 891
 QY 844 ttaaacccagatgtgtacagcttccagaaatagagagatgagcgagaaatggaag 902
 Db 892 ataaaccccttactcacaagcttgagaaatagatgaaggtgacctgagaaagct 950
 RESULT 14
 ABK16620
 ID ABK16620 standard; cDNA; 978 BP.
 AC ABK16620;
 XX
 DT 14-MAR-2002 (first entry)
 XX
 DE Human G-coupled receptor (GPCR) cDNA, seq ID No 29.
 XX
 KW Human: cytosolic; neuroprotective; immunosuppressant; mototropic;
 KW anti-inflammatory; anti-viral; gastrointestinal; cardiovascular;
 KW cerebroprotective; G-coupled receptor; cell proliferative disease;
 KW lymphoma; leukaemia; breast cancer; cirrhosis; neurological disorder;
 KW stroke; Alzheimer's disease; multiple sclerosis; mental retardation;
 KW cardiovascular disease; atherosclerosis; angina pectoris; indigestion;
 KW congestive heart failure; gastrointestinal disorder; dysphagia; AIDS;
 KW gastritis; autoimmune disorder; inflammatory disorder; Crohn's disease;
 KW systemic lupus erythematosus; metabolic disorder; diabetes; obesity;
 KW viral infection; herpesvirus; parvovirus;
 KW acquired immune deficiency syndrome; ss.
 XX

OS Homo sapiens.
 XX
 PN W0200190359-A2.
 XX
 PD 29-NOV-2001.
 XX
 PF 22-MAY-2001; 2001WO-US1683.
 XX
 PR 22-MAY-2000; 2000US-206222P.
 PR 25-MAY-2000; 2000US-207476P.
 PR 02-JUN-2000; 2000US-208634P.
 PR 02-JUN-2000; 2000US-208616P.
 PR 07-JUN-2000; 2000US-209868P.
 XX
 PA (INCYTE) INCYTE GENOMICS INC.
 XX
 PI Patterson C, Tribouley CM, Yao MG, Griffin JA, Thornton M, Lu Y;
 PI Kallik DA, Gandhi AR, Au-Young J;
 XX
 DR WPI: 2002-106199/14.
 DR P-PSDB: AA080498.
 XX
 PS Claim 5; Page 140; 148pp: English.
 XX
 CC The invention relates to a novel human G-coupled receptor (1). (1) and
 CC its corresponding polynucleotides are useful for diagnosing, treating or
 CC preventing cell proliferative diseases (e.g. lymphoma, leukaemia, breast
 CC cancer or cirrhosis), neurological disorders (e.g. stroke, Alzheimer's
 CC disease, multiple sclerosis or mental retardation), cardiovascular
 CC diseases (e.g. atherosclerosis, angina pectoris or congestive heart
 CC failure), gastrointestinal disorders (e.g. dysphagia, indigestion or
 CC gastritis), autoimmune/inflammatory disorders (e.g. AIDS, Crohn's disease
 CC or systemic lupus erythematosus) or metabolic disorders (e.g. diabetes
 CC or obesity), or viral infections (e.g. infection by herpesvirus or
 CC parvovirus). ABK16615-ABK16637 represent novel human G-coupled
 CC receptor coding sequences of the invention.
 CC
 SQ Sequence 978 BP; 228 A; 258 C; 173 G; 319 T; 0 other;
 CC
 Query Match 27.7%; Score 255.8; DB 24; Length 978;
 Best Local Similarity 55.3%; Pred. No. 4,4e-69;
 Matches 497; Conservative 0; Mismatches 402; Indels 0; Gaps 0;
 QY 4 aatcagaagctgtgaactagatgatacttattctgggctcaccaaaagctgaacctcag 63
 Db 52 aaccaaaccacatccagtaattcatctcccgaggattttcaagaagatgagatcaca 111
 QY 64 ggaattatctctctcttcttcttctcattgtctatctgtgtcttcttcggacaatgctc 123
 Db 112 aacctctctctgtgtcttcttcttctgtatgtacccgtgacatgattggaagcgctc 171
 QY 124 atcatcatcgaanaatctatagcaacacctgtgacagcgccatgatatgttctctctg 183
 Db 172 atcatgtgtcgtacagcttgatgatacgttcaatcccatcagcccatctctcttcgcc 231
 QY 184 acactgtgctgtgagacatctgcacaaagcatatccgaaagtgtctggagcc 243
 Db 232 aatctatctcttgatgatacttccatcttccaaactcagctcccaaatgctgtgagat 291
 QY 244 atgctacatcgaanaatccattcatatgcaagctgcatgcccagctctctgttc 303
 Db 292 attcaacaaagatcaatccatctcttatgagagctgatacaacagatgtactttct 351
 QY 304 acatgtctctgagagctgagatggtctctcccaacatgagctatgacagctatg 363
 Db 352 atgtgtctgtcgtcatctgacaattgctcttgaggagcatggtcatgaccacttg 411
 QY 364 gccattgttccctcttcatcagtaactatgatgaaccacatgtgtgtagcctg 423

Wed Jun 26 09:41:36 2002

us-09-975-308-8.rng

Page 15

Search completed: June 21, 2002, 12:29:40
Job time: 3865 sec

THIS PAGE BLANK (USPTO)

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

No.	Score	Query	Match	Length	DB	ID	Description
1	199.8	21.6	900	4	US-09-085-371-5	Sequence 5, App1	
2	189.6	20.5	966	3	US-08-748-506-7	Sequence 7, App1	
3	184	19.9	966	3	US-08-748-506-8	Sequence 8, App1	
4	182.8	19.8	966	3	US-08-748-506-5	Sequence 5, App1	
5	182.2	19.7	1290	2	US-08-827-291A-1	Sequence 1, App1	
6	177	19.2	966	3	US-08-748-506-6	Sequence 6, App1	
7	174	18.8	1713	2	US-08-467-948A-1	Sequence 1, App1	
8	174	18.8	1713	3	US-08-467-947A-1	Sequence 1, App1	
9	149	16.1	984	3	US-08-748-506-9	Sequence 9, App1	
10	94	10.2	1828	3	US-08-488-876-2	Sequence 2, App1	
11	86.2	9.3	963	4	US-09-439-313-526	Sequence 526, App1	
12	83	9.0	1474	1	US-08-465-980-1	Sequence 1, App1	
13	83	9.0	1474	2	US-09-053-303-1	Sequence 1, App1	
14	83	9.0	1474	2	US-09-053-303-1	Sequence 1, App1	
15	55	6.0	1330	1	US-08-599-252-84	Sequence 84, App1	
16	55	6.0	1330	1	US-08-436-074-57	Sequence 57, App1	
17	55	6.0	1320	5	PCT-US86-0635F-84	Sequence 84, App1	
18	55	6.0	1320	5	PCT-US86-0635F-84	Sequence 84, App1	
19	52	5.6	1506	1	US-07-937-609-13	Sequence 13, App1	
20	52	5.6	1506	4	US-08-029-170-13	Sequence 13, App1	
21	51.4	5.6	7218	1	US-08-332-463-14	Sequence 14, App1	
22	46.2	5.0	1336	1	US-07-578-892A-4	Sequence 4, App1	
23	46.2	5.0	1969	1	US-07-937-609-28	Sequence 28, App1	
24	46.2	5.0	1969	4	US-08-029-170-28	Sequence 28, App1	
25	45.6	4.9	1621	1	US-07-937-609-22	Sequence 22, App1	
26	45.6	4.9	1621	4	US-08-029-170-22	Sequence 22, App1	
27	45.6	4.9	1666	4	US-08-029-170-30	Sequence 30, App1	

28	44.2	4.8	2243	1	US-07-937-609-15	Sequence 15, Appl
29	44.2	4.8	2243	1	US-08-029-170-15	Sequence 15, Appl
30	44.2	4.8	2243	1	US-07-978-892-15	Sequence 1, Appl
31	43	4.7	1440	1	US-07-937-609-25	Sequence 25, Appl
32	43	4.7	2015	4	US-08-029-170-25	Sequence 25, Appl
33	42.2	4.6	1209	5	PCT-US85-05616-1	Sequence 1, Appl
34	41.6	4.5	936	2	US-08-288-663A-13	Sequence 13, Appl
35	41.6	4.5	1194	2	US-08-288-663A-2	Sequence 2, Appl
36	41.6	4.5	1288	2	US-08-288-663A-3	Sequence 3, Appl
37	41.4	4.5	1260	1	US-07-866-979-3	Sequence 3, Appl
38	41.4	4.5	1260	2	US-08-466-906B-3	Sequence 3, Appl
39	41.4	4.5	1260	3	US-08-706-281A-3	Sequence 3, Appl
40	41.4	4.5	1260	4	US-09-701-746-3	Sequence 3, Appl
41	41.4	4.5	1260	4	US-09-097-231-3	Sequence 3, Appl
42	41.4	4.5	1621	1	US-08-722-001-13	Sequence 13, Appl
43	41.4	4.5	1776	1	US-08-322-001-29	Sequence 29, Appl
44	41.4	4.5	2140	1	US-08-334-698-1	Sequence 1, Appl
45	41.4	4.5	2140	1	US-08-328-937-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
TC-00-085

```

Sequence 5, Application US/09085371
Patent No. 6218358
GENERAL INFORMATION:
APPLICANT: Firestein, Stuart
APPLICANT: Zhao, Haqiang
TITLE OF INVENTION: Functional Expression of, and Assay for, Functional Cellular
FILE REFERENCE: P01511US2 / 09805059
CURRENT APPLICATION NUMBER: US/09/085,371
CURRENT FILING DATE: 1998-05-19
PRIORITY APPLICATION NUMBER: US 08/891,243
PRIORITY FILING DATE: 1997-07-10
PRIORITY APPLICATION NUMBER: US 60/045,961
PRIORITY FILING DATE: 1997-05-07
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn version 3.0
SEQ ID NO 5
LENGTH: 900
TYPE: DNA
ORGANISM: Rat
US-09-085-371-5

```

Query Match	21.6%;	Score 199.8;	DB 4;	Length 900;
-------------	--------	--------------	-------	-------------

Matches 460; Conservative 0; Mismatches 397; Indels 12; Gaps 1,

QY	16	gtacctgattcattatctcttgggcctccaaacaaagctctgaaccccgagatattcttc	75
Db	28	gtgagtgaaatttgctgctgggtcttcocagctccctgcgccacctgagctactactt	87
QY	76	ctctctttctccattgtctacgtctggtcttcctgcgcaacatgcctcatcatctgccc	135
Db	88	tctctctctctctgcgactatgctgtgtgtgtacgtgaacacatgcctcatcatatagca	147
QY	136	aaaatctatagcaaacctctgctacagcccatatgattgtctctctcctgaacctgctgt	195
Db	148	attggaaccaccccaacccctcccaaaaacccatgattctttctctgctcatatagtcat	207
QY	196	gttgacatcattctgcacacaagaatcatatacgaagaatgc-----tggggacc	243
Db	208	ctggagagatttgtaatgcactcgtctacagatctcctaagaatgcctgctgctcatatgtccc	267
QY	244	atgttaacatcagaacaaataacattctatgacagctgcacatgtcccgctctctctgttc	303
Db	268	aaggaagaacacatgycacagctgattctccctttgaggacagacagaacacaaactctaatcttc	327
QY	304	acaaagtctctctggagctgagatgctctctctccacacatgycctatgacagctaatgtg	363

```

Db 328 ctgggcttggtgcaagaagtgctctctcttgctggtacatgacgcctatg 387
Oy 364 gcaattgttccctcttaatacagtaactatataagacacacatggtgtagcctg 423
Db 388 gcatctgtcatccactccactaccccgctcatgtgtagcgcggtagtgtagcagag 447
Oy 424 ctaagcatgtcatgtgctatgtgagcaacaaattcctggtgtagacacagctatcatg 483
Db 448 gcaagctgagatccggtcgagaggtttgtatctcatatggttaaaatttctctattct 507
Oy 484 aggttgatcttctgtgggccaacacacatgacacacttctctgtgagatacccatg 543
Db 508 cgcctgtcttacttggtcccaacacacatcacttctctgtgtagtggtctcattg 567
Oy 544 ctgctgttgcctgtagcctctgtaagaataatgagtgtagtgtagtgtagtatt 603
Db 568 ctcaactgttcaatgcatgacatgccaagcagagcttcaagacttgcctgagcatt 627
Oy 604 accctggcacaagggaacttattcttcaactgcatctcctcctatggtttatcattgtgct 663
Db 628 ttattctgtcggaccgcctctctgtcatcgtgggacatccatcagcacaaggtgct 687
Oy 664 attccgtatccgcagacagtagaagaagcagaagagccttccacatgctcatcctcat 723
Db 688 gtgatgcgcatccctcctcgtcgtcgtcgccacataaagccttcaacctgtgctccac 747
Oy 724 ctcaagtggtgaccccttacttacttctcctgtaactacacactacacgcctcctgtcc 783
Db 748 ctcaactgttgatcatcttctatgacagcagatatttcaatcatgcaagcctaagca 807
Oy 784 agctatacttgaagaagacagtgtagtgcactctactactctgtgacccacaca 843
Db 808 ctctcagcttttgacacacacacagctgctcctgtactcactcagctgtcatgtgacgt 867
Oy 844 ttaaacccgagtggtgtacagcttccagaa 872
Db 868 ttaacccaatcactactgcttgtagca 896

```

```

RESULT 2
US-08-748-506-7
Sequence 7, Application US/08748506
Patent No. 6159707
GENERAL INFORMATION:
APPLICANT: Romnett et al.
TITLE OF INVENTION: NOVEL SPERM RECEPTORS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leydig, Volt & Mayer, Ltd.
STREET: Two Prudential Plaza, Suite 4900
CITY: Chicago
STATE: IL
COUNTRY: US
ZIP: 60601-6780
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/748,506
FILING DATE: 08-NOV-1996
CLASSIFICATION: 435
PRIORITY APPLICATION: DATA:
APPLICATION NUMBER: US 60/033,751
FILING DATE: 09-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 74940
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-616-5600
TELEFAX: 312-616-5700

```

```

; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 966 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-748-506-7

```

```

Query Match 20.5%; Score 189.6; DB 3; Length 966;
Best Local Similarity 51.8%; Pred. No. 5.3e-50;
Matches 429; Conservative 0; Mismatches 399; Indels 0; Gaps 0;

```

```

Oy 71 tcttctcttcttctctctctctctctctctctctctctctctctctctctctctctca 130
Db 98 tcttctctctctctctctctctctctctctctctctctctctctctctctctctctca 157
Oy 131 ttgcacaaatctatagcaaacaccttgatgaagcccatatgattgttctctctgacatg 190
Db 158 ttgcattttgtacacagctccatcttctacacaccccatgcttcttctgacacattgt 217
Oy 191 ctgttgtagactatctctgcaacaacacatcatatcgaagatgctgggacacatgctaa 250
Db 218 ctctctctgagattggtcttacttctctctctctctctctctctctctctctctctg 277
Oy 251 catcagaataatccatcttcaataatgacgtgcatgctccagctctctctctctctct 310
Db 278 gtgagcccgagagagctcttggagaggttgccctacacagatggtcttctctctct 337
Oy 311 ctctggagctgagatggttctctctctcaacacatgacatgacatgacatgacatg 370
Db 338 ttgcatttactgagctgctctctctctctctctctctctctctctctctctctct 397
Oy 371 gttccctctctctctctctctctctctctctctctctctctctctctctctctctct 430
Db 398 gtccccacacacacacacacacacacacacacacacacacacacacacacacacac 457
Oy 431 tggatcgtctatctgacatcaccatctcctggttgacacacacacacacacacacac 490
Db 458 tctcatggtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 517
Oy 491 ctctctctctctctctctctctctctctctctctctctctctctctctctctctctca 550
Db 518 acttctctctctctctctctctctctctctctctctctctctctctctctctctct 577
Oy 551 tgcctgttagcctctgtaagataatgagtgtagtgtagtgtagtgtagtgtagtgtag 610
Db 578 ttgctctgtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 637
Oy 611 ccataaggagacttatctctctctctctctctctctctctctctctctctctctct 670
Db 638 tatttatgcatcttctctctctctctctctctctctctctctctctctctctctct 697
Oy 671 gtaacgacagtagaagaagcagaagagccttctcaacatctcatctcatcacaag 730
Db 698 tgatgctcttctctctctctctctctctctctctctctctctctctctctctctct 757
Oy 731 tggtagccttactatctctctctctctctctctctctctctctctctctctctctct 790
Db 758 tagtcacacttctctctctctctctctctctctctctctctctctctctctctct 817
Oy 791 cattgaaagacagagtgtagtgcactctcatctctctctctctctctctctctctca 850
Db 818 caccagagtgagacaaacttctgcccctctctctctctctctctctctctctctctct 877
Oy 851 cgatggtgtacagcttccagacatagagagatgacagagagatagga 898
Db 878 ccattatctacagctttaaagaaacaaagaaagaaagaaagaaagaaagaaagaa 925

```

```

RESULT 3
US-08-748-506-8

```

```

: Sequence 8, Application US/08748506
: Patent No. 6159707
: GENERAL INFORMATION:
: APPLICANT: Ronnett et al.
: TITLE OF INVENTION: NOVEL SPERM RECEPTORS
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Leydig, Volt & Mayer, Ltd.
: STREET: Two Prudential Plaza, Suite 4900
: CITY: Chicago
: STATE: IL
: COUNTRY: US
: ZIP: 60601-6780
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/748,506
: FILING DATE: 08-NOV-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/033,751
: FILING DATE: 09-NOV-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: REFERENCE/DOCKET NUMBER: 74940
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312-616-5600
: TELEFAX: 312-616-5700
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 966 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-748-506-8

```

```

Query Match          19.9%; Score 184; DB 3; Length 966;
Best Local Similarity 51.1%; Pred. No. 3.1e-48;
Matches 433; Conservative 0; Mismatches 415; Indels 0; Gaps 0;

```

```

QY 491 cttctgtggtggaacaccattgaccattctctgtgtgagatgaccccatgctgctt 550
    ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
Db 518 ACTTGTGTGAGACCTGTGAGATGAGACCACTTCTTGTGTGATCTTCCACCTCTCTG6CAC 577
QY 551 tctcccttgagccctgtaagaatcatcatgagtgatggtgtgtgtgtgtgtgtgtgtgt 610
    ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
Db 578 TTGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 637
QY 611 ccataggagacttattcttaccctgcatctcctctagtgtttalcatlctgtgtcatctcc 670
    ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
Db 638 TATCTAGCCCTTTTGTGCTGATCATTTATTTATGTCAGAAATTCGTGGCAGTGTG 697
QY 671 gtaicgcacagtagaaggaaggaaggaagccttcacacatgcatctcatctcaag 730
    ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
Db 698 TGAATGCTTACCTGAGGGGCGCCACAAAGCCCTTTCACCTGTTCTCCACCTACTTG 757
QY 731 tggtagacccttactatctctcgtlaatcatcatataccgcccctgtctcagata 790
    ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
Db 758 TAGTCACACTCTTTTATGCTGAGATCTGTACCTATTGAGGCTTAAGTCTAGCCACT 817
QY 791 cattgaagaagacaggtggtgagctgcaactcatatctcttggactcccaatlaacc 850
    ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
Db 818 CACACAGGAATGACAAACTCTTGCCCTCTTACACACAGCAGATGACATCCATGTGAACC 877
QY 851 cgatggtgacagctcccaagatagagatgcaagcaggaatgaaggtgttgcac 910
    ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
Db 878 CTATCATCTATAGTTTAAGGAACAGATGTCAAGGCACACTGAGAAGATTTCTGCCC 937
QY 911 ttctgaa 918
    |||||
Db 938 TGAATAAA 945

```

```

RESULT 4
US-08-748-506-5
: Sequence 5, Application US/08748506
: Patent No. 6159707
: GENERAL INFORMATION:
: APPLICANT: Ronnett et al.
: TITLE OF INVENTION: NOVEL SPERM RECEPTORS
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Leydig, Volt & Mayer, Ltd.
: STREET: Two Prudential Plaza, Suite 4900
: CITY: Chicago
: STATE: IL
: COUNTRY: US
: ZIP: 60601-6780
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/748,506
: FILING DATE: 08-NOV-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/033,751
: FILING DATE: 09-NOV-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: REFERENCE/DOCKET NUMBER: 74940
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312-616-5700
: TELEFAX: 312-616-5600
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 966 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)

```

US-08-748-506-5

Query Match 19.8%; Score 182.8; DB 3; Length 966;
 Best Local Similarity 51.2%; Pred. No. 7.4e-48;
 Matches 424; Conservative 0; Mismatches 404; Indels 0; Gaps 0;

```

QY 71 tcttctcttcttctcattgtctatcttctgtgcttctctcggaactgtctcatca 130
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 98 tcttcaaccttcaaccttctcattgtctatcttctgtgcttctctcggaactgtctcatca 157
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 131 ttgcaaaatcatagcaaaccttgcataagcccatgattgttctcttctgcacac 190
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 158 ttgcaaaatcatagcaaaccttgcataagcccatgattgttctcttctgcacac 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 191 ctgttgtagacatctctgcacaaagaatcatatcgaagatgtctgggaaccatgctaa 250
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 218 ctgttgtagacatctctgcacaaagaatcatatcgaagatgtctgggaaccatgctaa 277
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 251 catcagaataatcatcttcatatgcagagctgcatgtccagctctcttctcatatgct 310
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 278 gtgagcccaagagatctctcggagagatgtgccaacacagatgttcttctcgcatcttt 337
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 311 ctctggagagctgagatgtctctcttccacacatgagctcatgacgctatgtggcattt 370
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 338 ttggtatfaactgagctgctgttggcagccatggcctttgacccctgcatgagccatgat 397
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 371 gtttccctcttcatcagctacattatgaacacacatgctgtgtgagcttgcacga 430
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 398 gcttccctcttcatcagctacattatgaacacacatgctgtgtgagcttgcacga 457
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 431 tggtagatgctatgcagctacacatctcctgggtgcacacagctcttcatcagatgctaa 490
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 458 ttgtcatgagggaatggagatgcatgataagctgggacaaacaaattttatgttctctgta 517
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 491 cttctgtgtggccaacacacacattgacacatctctgtgtgagataccccaattgtgctt 550
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 518 acttctgtgtggccaacacacacattgacacatctctgtgtgagataccccaattgtgctt 577
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 551 tgtctgttagccctgttagaatacaatgagatgagatgtgtgtgtgtgtgtgtgtgtgtgt 610
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 578 ttgctgtgttagccctgttagaatacaatgagatgagatgtgtgtgtgtgtgtgtgtgtgt 637
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 611 ccatagggagcttattcttcttaccgcatcctcctcctcctcctcctcctcctcctcctc 670
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 638 tatctagcccaattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 697
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 671 gtatccgacagtagaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 730
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 698 tgatgcttcaacctgagagggggcccatgaaagcttcttccacctgtgtgtgtgtgtgtgt 757
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 711 tggtagacatttcttctcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 790
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 758 tagtcaacatttttttttttttttttttttttttttttttttttttttttttttttttttt 817
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 791 catttgaagagagaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 850
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 818 caccaggaatgacaaattcttggcccttcttaccagatgtagcatcagcatcagcatcagcat 877
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 851 cgaatgtgtacagcttccagaatgaggaatgaggaatgaggaatgaggaatgaggaatgagga 898
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 878 ctatcatctatagtttaaggaacaaaggaatgaggaatgaggaatgaggaatgaggaatgagga 925
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 5
 US-08-827-291A-1
 ; Sequence 1, Application US/08827291A
 ; Patent No. 5874243
 ; GENERAL INFORMATION:
 ; APPLICANT: Macina, Roberto
 ; APPLICANT: Sathie, Ganesh
 ; TITLE OF INVENTION: NOVEL OLRCIS RECEPTOR
 ; NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:
 ADDRESSEE: SmithKline Beecham Corporation
 STREET: 709 Swedeland Road
 CITY: King of Prussia
 STATE: PA
 COUNTRY: PA
 ZIP: 19406

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/827,291A
 FILING DATE: 28-MAR-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:

ATTORNEY/AGENT INFORMATION:
 NAME: King, William T.
 REGISTRATION NUMBER: 30,954
 REFERENCE/DOCKET NUMBER: GP50001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 610-270-5015
 TELEFAX: 610-270-5090

TELEX:
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1290 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA

US-08-827-291A-1

Query Match 19.7%; Score 182.2; DB 2; Length 1290;
 Best Local Similarity 50.1%; Pred. No. 1.3e-47;
 Matches 454; Conservative 0; Mismatches 453; Indels 0; Gaps 0;

```

QY 3 gaacacagcgtgtgtagcagctatcttctgtgctcaccacaaagccctgaactca 62
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 307 gaattgaccttcaacctcgcagcttcttctggaatcttcaacacagcccccaccca 366
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 63 gggaaattctctcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 367 cacttctcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 123 catcatcatcgcacaaatctatagcaaaccttgcataagcccatgatagtattcttctct 182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 427 catggttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 486
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 183 gacatggtcgtgtgtgacatcatctgcacaaagcatcatcaggaatgtctggagac 242
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 487 ccaactgtccctcattgacacatcattctatgagcagcagcagcagcagcagcagcagcagc 546
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 243 catgtacatcagaataatcattcataltgagagctgtgacatgtccagctcttctgtt 302
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 547 ctacctgtcttgacgacaaagcattttctatggcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 606
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 303 cacatgtctctgtgagctgagatgtctcttccacacatgtggcattgacacacacacacac 362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 607 tacatcacctgttgcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 666
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 363 ggcatttgttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 422
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 667 tgcatttggccaccccttcaatcatcaccatcatcagccctaaattgtgtgaccttat 726
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 423 gctcagcatgtcatgtcatgtcatgtcatgtcatgtcatgtcatgtcatgtcatgtcatgtcat 482
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 727 gacttcccttttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 786
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


TELEFAX: 202-371-2540
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1713 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: both
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 116..1003
: US-08-467-947A-1

Query Match 18.8%; Score 174; DB 3; Length 1713;
Best Local Similarity 51.0%; Pred. No. 5.9e-45;
Matches 411; Conservative 0; Mismatches 395; Indels 0; Gaps 0;

QY 4 aatcacagcgtgtaacagtgatctatcttgggcccacacaaagcctgaactcag 63
DB 137 AATCAGACAAATGCTACAGAGTTCTCTACTGGGATTTCCTGGCCCAAGATTGAG 196
Y 64 ggaattactctctcttcttctcattctgcttctgcttctcgcgaacatgctc 123
DB 197 ATGCTCCTCTTGGGCTCTTCCCTGTTTATGTTTCACCCCTGCTGGGAATGGGACC 256
QY 124 atcatcatgcccaaaatctatagaacaccccttgcatcccaatgatagtttcctctg 183
DB 257 ATCTGGGGGCTCATCTCACTGGAGTCCAGACTCCACACCCCATCTTCTCTCTCA 316
QY 184 acactggtgtgttggagacatctgcacaaagaacatcatcagaagtctggagacc 243
DB 317 CACCTGGCCGCTGTCACATGCTCTATGCTGCAACACAGTGCCTCAATGCTGTGAC 376
QY 244 atgctaacatagaanaatacatctcatatgcaagctgcatgtccagctctctgttc 303
DB 377 CTCTGATCCAGCCAGCAACCCATCTCTTGTGCTGATGACACACTGACTTCTCTTT 436
QY 304 acatggtctcggagctgagatggttctcttcaacacacagccttgaagcgcctatgt 363
DB 437 TTGAGTTTGGCAATACATGCAATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 496
QY 364 gccattgttccctctctcatctacagtaactatctgaacacacatgctgtagcctg 423
DB 497 GCCATCTGCCACCTCTCTCCGATTTTCAATCATGACCTGGAAAGTGTGATCATCTGTG 556
QY 424 ctacagcatgcatagtgatgtcagtcacacattcctggtggtgacacagccttcatcag 483
DB 557 GGCATCACTTCTGAGACATGTGGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 616
Y 484 aggttgaattctgttgggccaacacacattgaacacattcctctgtgagatacccccat 543
DB 617 AGACTGCCCCCTTTTGTGGGCTGCTGTAATCAACCACTTCTTGTGAATCTGTCTGTC 676
QY 544 ctgagcttgcctgtagccctgtaagaataatgagtgatgtgtatgttgcataatt 603
DB 677 CTGAGGCTGGGCTGTGCTGATACCTGCTCAACAGGTGCTGATCTTTGAAGCTGCTGAT 736
QY 604 accctggcacaagaagaccttattcttcaactcgtcatctcctatggttttcatcattgt 663
DB 737 TTTCATCTCTGCTGGGACCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 796
QY 664 atttcgctatccgcaacagtgtaagaagaagagccttcaactgtctcatctat 723
DB 797 ATCTGAGGATCCAGTCTGGGGAGGCGCAGAAAGGCTTCTCCACTGCTCTCCAC 856
QY 724 ctacagtggtgacaccttactatctctctttaaactcaacattatcgccgtctctc 783
DB 857 CTGCGGAGTAGGAGACTTCTTTGGGAGGCGCCATCTGATGATGATGATGATGATGATGATG 916
QY 784 agctatcatcttgaagaagacaagt 809
DB 917 CGCATCTCTGAGGAGCAGCAAGAGT 942

RESULT 9
US-08-748-506-9
: Sequence 9, Application US/08748506
: Patent No. 6159707
: GENERAL INFORMATION:
: APPLICANT: Ronnett et al.
: TITLE OF INVENTION: NOVEL SPERM RECEPTORS
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Leydig, Volt & Mayer, Ltd.
: STREET: Two Prudential Plaza, Suite 4900
: CITY: Chicago
: STATE: IL
: COUNTRY: US
: ZIP: 60601-6780
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/748,506
: FILING DATE: 08-NOV-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/033,751
: FILING DATE: 09-NOV-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION: 74940
: REFERENCE/DOCKET NUMBER:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312-616-5700
: TELEFAX: 312-616-5700
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 984 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-748-506-9

Query Match 16.1%; Score 149; DB 3; Length 984;
Best Local Similarity 47.8%; Pred. No. 3.3e-37;
Matches 431; Conservative 0; Mismatches 470; Indels 0; Gaps 0;

QY 1 atgaatcacagcgtgtgtaacagtgatctatcttctgggcccacacaaagcctgaactc 60
DB 28 AGAATGAGGACTTGTGCTCTGAGTTTCACTTGAAGGGTACCTGTGGCCGAGCACTTG 87
QY 61 caggaaattactctctcttcttctcattctgtcttcttctcgcgaacatg 120
DB 88 AAGATCTCTTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 147
QY 121 ctcatcatcattgcgaanaatacatctatagaacacaccttgcatcagccatgatagtttcct 180
DB 148 CTCATTAATTACCATCACTGCTGAGCACCGACCTGACAGACGCCATGACTTCTTCTC 207
QY 181 ctgacacggtcgtgtgtgacatctcgtcacaaagaacatcatcgaagaagtgctgggg 240
DB 208 AGCACCTTCTCTTGTGTGAGGTGTTTATTAATCTACTGCTATCCCAAGCTCTCTCAC 267
QY 241 accatgctaacatcagaanaatacatctcatatgagcgtgcatgtlccagctctctctg 300
DB 268 ATCATTTCTGAGGAGGCAAAAGATTCCTTTGGGGTCTGCTTCTCAAGAGCTTCTGTC 327
QY 301 ttcaatggtctcgtggagctgagatggttctcttcaacacacagcgtctatgacgcctat 360
DB 328 TATCTTGTCTGGGGGCAACAGGTTTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 387

Oy		361	gtggccatttglttccccttcctcctcaatgaagtaactataatgaaccacatttgtytagcc	42
Db		388	ctfgcccatctgcAAACCTCtCAATTTCCAACCATATGAGCCCAAGAtgTTCCTT	447
Oy		421	tgtcgagaatgtaatgatgatgcataatgcagtacaattccctgggtgacaaagctttatc	480
Db		448	ctcgtTACTGTCTGTATTATTGGGGCTTCCTTCAATGCGCACGTCAGTGtGATGCTT	507
Oy		481	atgaagtgtagcttcgtgaggccaacacacatgagcacctcttcctgtgtagatacccca	540
Db		508	TCCAAGACATTTTACTGTGGGCCAACAATTATTCCTCATTtTCTGTGAATTTGGACCA	567
Oy		541	tgtctgagcttgtctcgttagacctgtaaataatgaatgaatgtagtgytlatgtgtcyat	600
Db		568	CTGGCAAACTCTCTCTGTTCCAGAACCAGGCTATTTGAGATGCCTGTTTTAACCTGCT	627
Oy		601	atlaacctgagcataaggagacatttatcttaacctgcatactccctatggtttatcatgtt	666
Db		628	GTAATTGtGCTTTTGGCTTCCTTCTTATAACCATCTTTGGATACAGCAATAATAgTgTC	687
Oy		661	gtctattccgtatcccgacaggtgaaggagcaaggagagccctctaactgctcatct	720
Db		688	ACCAATAGGAACATCCCTTCCAGCCAGGAGGAGAGAGAGCTTTTCCACCTGCTCTCT	747
Oy		721	catecaagatggtgagcccttactatctctcgtlaactciaaacctatatccgccctgct	784
Db		748	CATCTCATTTGTCCTCTCTCTAATGATGAGCAGCTGTGCAATTAAATACCTGAACCAAG	807
Oy		761	tccagctatacttgaaagagcaagtgtagctgactctatacttactctgtgactccc	840
Db		808	CAGAGAACAGAGTGtGACCAACAGAGAGAGGCTGCTCTTGTGMACATGTGTTGACACCC	867
Oy		841	acattaacccgatggtgtacagcttcagaataggaagatgagcaggaattagaag	900
Db		866	CTTCTGAACCTGTCTATCTACACCTCGGCAACAGCAGTCCACAGGCTCTCAGGAT	927
Oy		901	g 901	
Db		928	G 928	
	RESULT	10		
	US-08-988-876-2			
	; Sequence 2, Application US/08988876			
	; Patent No. 6063596			
	GENERAL INFORMATION:			
	APPLICANT: Lal, Preeti			
	APPLICANT: Bandman, Olga			
	APPLICANT: Hillman, Jennifer L.			
	APPLICANT: Yue, Henry			
	TITLE OF INVENTION: G PROTEIN COUPLED RECEPTORS ASSOCIATED			
	TITLE OF INVENTION: WITH IMMUNE RESPONSE			
	NUMBER OF SEQUENCES: 9			
	CORRESPONDENCE ADDRESSES:			
	ADDRESSEE: Incyte Pharmaceuticals, Inc.,			
	STREET: 3174 Porter Drive			
	CITY: Palo Alto			
	STATE: CA			
	COUNTRY: USA			
	ZIP: 94304			
	COMPUTER READABLE FORM:			
	MEDIUM TYPE: Diskette			
	COMPUTER: IBM compatible			
	OPERATING SYSTEM: DOS			
	SOFTWARE: FASTSEQ for Windows Version 2.0			
	CURRENT APPLICATION DATA:			
	APPLICATION NUMBER: US/08/988,876			
	FILING DATE: Herewith			
	CLASSIFICATION:			
	PRIOR APPLICATION DATA:			
	APPLICATION NUMBER:			
	FILING DATE:			
	ATTORNEY/AGENT INFORMATION:			

NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0441 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1828 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PROSNOTO1
CLONE: 364702
US-08-988-876-2

Query Match	10.2%	Score 94	DB 3	Length 1828
Best Local Similarity	48.2%	Pred. No. 9.6e-20		
Matches	356	Conservative	0	Mismatches 375; Indels 7; Gaps 3
QY	3	gaatcagcgtgtgtaacgaggtgtcaatctatctcgtggccctcacaaaaagcctgaactcca	62	
Db	792	GAATCAAAAGAGATGCTTATATTTCTCTCTCTCTCAAGAAAGTCTCAGGGGATCCAGAACTGCA	851	
QY	63	gggaattaccttcctcctctttttctccatctgtctatctctgttggtcttcctggcaacttgct	122	
Db	852	GCCAGTCCCTTGCGGGGCTGTCTCTGCATAGTGCCTGGTGACACGGTCTGGGGAACTGGCT	911	
QY	123	catcatcatgtccaaaatctatagcaaacacctgtagccgaatgtagttctctct	182	
Db	912	CATCATCCCGGCCCATCAGCCCTAGCTCCACCTCCACCTCCACACCCCATGTACTCTTCTCTCC	971	
QY	183	gaacatgagctggtgtgagacatcatctgcaacaagaatcatatccgaagaatgctggggac	242	
Db	972	CAACCTGCTCTTGCTGCTTACATCGGTTTACCTCCACCTCCACACGAGTCCCAAGATGATATGGGA	1031	
QY	243	catgtcaacatcagaaaaataacatltcaatatgcaagctgcatgtccagctctcttggt	302	
Db	1032	CATCCAGATCTCACAGCAGAGATCATCTCTATMGACGGCTGGCTGACTGATGATGTCCTCTT	1091	
QY	303	caaatgtctctcggagagctgagatg----gttctctcacacaacatgacatgacgcgt	358	
Db	1092	TGCCATTTTGGAGGATGAGAGAGACATGCTCTCTGAGTGTATGGCCATATGACGTGGT	1151	
QY	359	atgtagccattgttccctctcattacagtaactatataagaccacataatgtagtag	418	
Db	1152	TTTGAGCCATGTGCACCCCGGTATATCATTCACCATATGAAACCGGTTTCTGTGCGCTT	1211	
QY	419	cctgtctcagcatgtgtaatgtctatgtcagtcacc--aatctcgggtgcacaagctct	476	
Db	1212	TCTAGTTTGTGTGCTTTTCTTTTTCACAGCTTTTAAAGCTCCACAGCTGCAACACTTGAT	1271	
QY	477	tataatgaggttagcttctcgtctgtagccctgtlaagaatcaatgagtgatgtatgttc	536	
Db	1272	TGCCCTTACAAAGTACGCTGCTCAAGAGATGTGGAATTTCTTAATTTCTGTGTGACCCCTTC	1331	
QY	537	cccatctgtcgtctgtctcctgtagccctgtlaagaatcaatgagtgatgtatgttc	596	
Db	1332	TCAACTCTCCCATCTGTGATGTGTGAGACCTTCACATTAACATTAACATCATATTTCCG	1391	
QY	597	tgataataccctgagcaataaggagacttaattctctaacctgcatctcctatggtttatcat	656	
Db	1392	TGCTGCCATATTTGGTTTCTTTCCTCATTCAGGAGACCTTTCTCTTAAGTAAAAATTCCT	1451	
QY	657	tggtctattctcgtatccgacagatgaaagagcaagagagcctctcaaatgctc	716	
Db	1452	TTCTCTCATCTGGAAGGTTTCTCATCTACAGTGTGGAAGATATAAA--CCTTCTCACCTGTGG	1510	
QY	717	atctcatctcaacagctggt	734	

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2002, 09:13:22 ; Search time 1694.03 Seconds

(without alignments)
7361.845 Million cell updates/sec

Title: US-09-975-308-8

Perfect score: 924
Sequence: 1 atgatacagcgctgtgaac.....ttgcattctcgaacactag 924

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estinu:*
5: em_estov:*
6: em_estopl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	323.2	35.0	470	12	AQ428256 CITR1-E1-
2	200.6	21.7	2021	11	BC016940 Homo sapi
3	188.6	20.4	580	12	AZ593814 IM0405K03
4	175.4	19.0	3063	11	AK016560 Mus muscu
5	173.8	18.8	1394	11	AK017036 Mus muscu
6	173.2	18.7	726	12	AZ255734 RPCI-23-1
7	172	18.6	616	12	BH342053 CH230-51L
8	171	18.5	762	10	BC193339 RST12467
9	170	18.4	698	12	BH069789 RPCI-24-3
10	164.6	17.8	611	12	AZ103967 RPCI-23-3
11	159.2	17.2	853	12	BH31857 CH230-46N
12	159	17.2	679	12	AG167722 Pan trogl
13	157.4	17.0	642	12	AZ696227 2M0241J24
14	156.6	16.9	600	12	AQ976178 RPCI-23-3
15	155.6	16.8	796	10	BC197640 RST17016
16	154.4	16.7	842	12	AZ725747 RPCI-24-9
17	153.4	16.6	699	12	AZ555190 RPCI-23-1

18	153.2	16.6	1501	11	AK016338 Mus muscu
19	152.6	16.5	612	12	AZ396764
20	152.4	16.5	805	12	BH272774 CH230-56L
21	150.2	16.3	629	12	AZ019257 RPCI-23-2
22	150	16.2	630	12	AQ503914 RPCI-11-2
23	149	16.1	605	12	AZ642411
24	148.8	16.1	794	12	AZ913406 RPCI-24-1
25	147.8	16.0	627	12	AQ240757
26	147.8	16.0	680	12	AZ709687
27	147.6	15.9	814	12	AZ699230 RPCI-23-2
28	147.2	15.9	686	12	AZ086625 RPCI-23-2
29	146.8	15.9	634	9	BB635510
30	146.4	15.8	542	12	AZ252110 RPCI-23-4
31	146	15.8	743	12	BH322680 CH230-197
32	145.6	15.8	752	12	BH065531 RPCI-24-2
33	144	15.6	632	12	AZ765752 IM0562E19
34	143.8	15.6	542	12	AQ632709 RPCI-11-4
35	143.8	15.6	634	12	AZ507680
36	143.2	15.5	635	12	AQ527378 RPCI-11-3
37	142.8	15.5	667	12	BH293359 CH230-92C
38	141	15.3	752	12	BH362510 CH230-127
39	140.2	15.2	545	12	AQ389818 RPCI11-14
40	140.2	15.2	632	12	AZ648626 IM0517B18
41	140.2	15.2	660	12	AZ380178 IM0136A04
42	140.2	15.2	732	10	BC6068751 H3068H10-
43	139.8	15.1	728	12	BH279965 CH230-38K
44	139.2	15.1	521	12	AZ407823 IM0178P09
45	138.6	15.0	594	12	AZ599283 IM0414H04

ALIGNMENTS

RESULT 1
LOCUS AQ428256
DEFINITION CITR1-E1-2578F11.TF CITR1-E1 Homo sapiens genomic clone 2578F11.
DNA sequence.
ACCESSION AQ428256
VERSION AQ428256.1 GI:4496022
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 470)
Zhuo,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and Venter,J.C.
Autors
Title
Use of BAC End Sequences from Caltech Libraries for Sequence-Ready Map Building
Unpublished (1997)
JOURNAL
COMMENT Other-GSS: CITR1-E1-2578F11.TF
Contact: Shaying, Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9112 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbest@igf.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

FEATURES

source
1..470
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2578F11"
/clone_lib="CITR1-E1"
/sex="male"
/cell_type="sperm"
/note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;

BASE COUNT 119 a 122 c 80 g 147 t 2 others
ORIGIN

Query Match 35.0%; Score 323.2; DB 12; Length 470;
Best Local Similarity 98.8%; Pred. No. 3.2e-72;
Matches 325; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 atgaacacagcgtgtgaactgaatcattatcttggcctcaacaaagcctgaactc 60
|||||
Db 142 ATGAATCAGAGCGCTTGAGTGCATATTCCTGGCCCTCAGCAAAAAGCTGAATC 201
QY 61 caagggaattactc 120
|||||
Db 202 CAGAGATATATCT 261
QY 121 ctatcatcattgcaaaatctatagaaacaccttgcatacgccatgtatgttctt 180
|||||
Db 262 CTATCTATCTGCGATATCTATTAACACACCTTGATGCGCCATGATGTTTCTT 321
QY 181 ctgacactggctgtgtgacatcatctgcacaaagcctaccgaagatgctggg 240
|||||
Db 322 CTGACACTGGCTGTGTGGACATCATCTGCACAAAGCATACCGAAGATGCTGGG 381
QY 241 accatgctaatcagaataatcattcattatgcagctgcacatgtccagctctct 300
|||||
Db 382 ACCATGCTACATCAGAAAATACCATTTCAATGTCAGGTCGATGCCAGCTCTCTTG 441
QY 301 ttccatgtctctgggagctgagatgt 329
|||||
Db 442 TTCACATGCTCTGGGAGCTGAGATGCT 470

RESULT 2
BC016940 2021 bp mRNA linear HTC 09-NOV-2001
LOCUS Homo sapiens, similar to olfactory receptor, family 2, subfamily A,
DEFINITION member 4, clone IMAGE:4424116, mRNA.
ACCESSION BC016940
VERSION BC016940.1 GI:16877381
KEYWORDS HTC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2021)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mdedpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

REMARK
COMMENT
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
Series: IRAC Plate: 27 Row: 1 Column: 22
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Genomescan gene prediction, Similarity but not identity
to protein

FEATURES
Location/Qualifiers
1..2021
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4424116"
/tissue_type="Kidney, hypernephroma"
/clone_lib="NIH_MGC_89"
/lab_host="DH10B"
/note="Vector: pCMV-Sport6"

BASE COUNT 470 a 546 c 427 g 578 t
ORIGIN

Query Match 21.7%; Score 200.6; DB 11; Length 2021;
Best Local Similarity 51.5%; Pred. No. 1.2e-40;
Matches 461; Conservative 0; Mismatches 434; Indels 0; Gaps 0;

QY 4 aatcaacagcgtgtgaactgaatcattatcttggcctcaacaaagcctgaactcag 63
|||||
Db 438 ATATCAGACAAATGGTGCAGAGATTCCTCTACCTGGGATTTCTGGGCCAAGATTGAG 497
QY 64 ggaattatc 123
|||||
Db 498 ATGCTCA 557
QY 124 atcatcatgccaataatctatagcaaaccttgcatacgccatgtatgttctctg 183
|||||
Db 558 ATCTGCGGGCTCATCTCACTGAGCTCCAGACTCCACCCCATTTACTTCTCTCTCA 617
QY 184 acactggtctgtgtggaactcattcgaacaaagcctacatcacgaagatgctggagcc 243
|||||
Db 618 CACCTGCGCGTGTCAACTACGCTTATGCTTGCCTGACACAGTCCCGAGATGCTGTGAAC 677
QY 244 atgctaacatcagaataatcattcattatgcagctgcacatgtccagctctctgttc 303
|||||
Db 678 CTCTCTGATTCACACCAAGCCATCTCTTGTGCTGCTGATCATAATGACCTTCTCTTT 737
QY 304 acatggtctctggagctgagatggtctctctcaccacacatgacctatgacctatgtg 363
|||||
Db 738 TTGAGTTTTCACATCTACTGATGCTCTCTGTTGTCTGATGCTCTGATGCTCTGATGCT 797
QY 364 gccattgttccctctctcattacagtaactatgaacacacatgtgtgtagcttg 423
|||||
Db 798 GCGATCTGCCACCCCTCTCCGATATTCATCATCATGACCTGGAATGCTGCATCACTCTTG 857
QY 424 ctgacatgtctatgctatcagtcacgaatcctctctggtgacacagctctatcatg 483
|||||
Db 858 GCCATCACTTCTGACATGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 917
QY 484 aggttgccttctgtgggccaacacacatgacacacttctctgtgagatacccatg 543
|||||
Db 918 AGACTGCGCTTTTGTGGCTCTGGAATACACACACTTCTTGTGAATACCTGTCTGTC 977
QY 544 ctggtcttgcctgttagccctgtgaagaatcaagagtgtagttagttagttagttagt 603
|||||
Db 978 CTCAGGCTGGCGCTGTGTGATACCTGTGCTCAACAGAGTGGTATCTTTCAGAGCTGTGATG 1037
QY 604 accctggccatagggagacttattcttaacctgcacccctatggtttatcatgtgtgtc 663
|||||
Db 1038 TTCAATCTGTGTGGACACACTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 1097
QY 664 attctcgtatccgacagtagtaaggaaggaaggaaggaaggaaggaaggaaggaagga 723
|||||
Db 1098 ATCTGAGAGATCCAGTCTGTGGGAGGCGCGAGAAAGGCTTCTTCACCTGTCTCCAC 1157
QY 724 ctcaacagtgtagcccttattctctctgtaaacattacacattatccgcctgtctcc 783
|||||
Db 1158 CTCTGCTGATGTGGACTCTTCTTGTGGAGCGCATCGATGATGAGGCGCTGAAGCTC 1217
QY 784 agctatcatcttgaagagacaaagtgtagtgcactatctatctgtgagcccaaca 843
|||||
Db 1218 CGCATCTCTGAGAGACAGAGAGAGAGTCTTCTTCTATTTTACGTTCTTTACACCCGATG 1277

QY 844 tttaaccgatggtgtacagcttcacgaataggagatgcaaggcaggaattagga 898
DB 1278 CTAACCCCTGATTACAACTGAGGAGATGTAAGGTCAAGGGTGCCCTGAGGA 1332

RESULT 3
AZ593814 580 bp DNA linear GSS 13-DEC-2000
LOCUS 1M0405K03R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION Clone UUGC1M0405K03 R, DNA sequence.
ACCESSION AZ593814
VERSION AZ593814.1 GI:11716004
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 580)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
'M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.,
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
JOURNAL Contact: Robert B. Weiss
COMMENT University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0405 row: K column: 03
Seq primer: CACACGAGAAACAGCTATCACC
Class: plasmid ends
High quality sequence stop: 580.
FEATURES
location/Qualifiers
1..580
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0405K03"
/clone.lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (g14732114|g14732114|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 114 a 162 c 129 g 175 t
ORIGIN

Query Match 20.4%; Score 188.6; DB 12; Length 580;
Best Local Similarity 59.4%; Pred. No. 1e-37;

Matches 320; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

QY 385 tacagtaactattgaaccacatgctgtgacctgtcagcatgcatgctatt 444
DB 1 TACAGCTCTAGGATGAGCCACAGATGTGGGGCCCTGGCCATGGATGTATGCTCATT 60

QY 445 gcaatcaccaatccctgggtggtgacacagctctatcatgaggtgtgacctgtgagcca 504
DB 61 AGTCTGTGTAATCATCTGTGACACTGGCCTGATGACAGGCTGTATCTGTGACCC 120

QY 505 aaacacattgacacattctctgtgataccccattgctggtctgtctgtagccct 564
DB 121 AAGTATCATCCACACTCTTCTGTGAGATTCCCCACTCCTCCTGCTCTGTACTCT 180

QY 565 gtaagaatcaatgaggtgagtgatgtgtgtgtcgtatataccctggtcgaataggacct 624
DB 181 ACATATGTAATATCCATTATGACTCTTTGGGAGATTCTTTTGGAGGCGCAATTTT 240

QY 625 attctaccctgacatccctgattgtttatcatgtgttcattctcgtatccgacagta 684
DB 241 GTCTTACCTTGTGATTCCTATGCTGCTGATGATTCACAGATCTGCGCATGCTTTGCT 300

QY 685 gaaggaagaagaagagcctctcaacatgctcatctcatcagtggtgaccttacc 744
DB 301 GAGGCGAAGAGAGAGGCGCTTTCTACCTGCTATCCACCTCATTTGGTCTGTGTAC 360

QY 745 tattctccgtatcttcaacctatctccgacctgtctcagcatatcattgaagaagac 804
DB 361 TACTCATCTGTGTTCTGTGTGCTGATGTCAGCCCTGCTTCCAGTCAAGTCCAGAAAC 420

QY 805 aaggtgtagctgactctatctatctgtgtgacccacataaaccgagtggtacagc 864
DB 421 AAAGTTACCTGATGTTATCTATGATCGTCAGGCCCAACCTCAACCCCTCATCTTATCA 480

QY 865 ttccagaataggagatgcaaggcagaataggaaaggtgttgcatttctgaacacta 923
DB 481 CTGAGGAGAACAGAGATGTCAAGCTTGCTGGCGAGATATTGGCCCTTCTTCACATTA 539

RESULT 4
AK016560 3063 bp mRNA linear HTC 19-JAN-2002
LOCUS Mus musculus adult male testis cDNA, RIKEN full-length enriched
DEFINITION library, clone:4932441H21: similar to T1 OLFACTORY RECEPTOR, full
insert sequence.
ACCESSION AK016560.1 GI:12855357
VERSION AK016560
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,
clone.lib:RIKEN full-length enriched mouse cDNA library
clone:4932441H21.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Carninci,P., and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (sites)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (sites)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Kono,H., Akiyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M.,
Sunii,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,

TITLE
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matakaki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE
PUBMED
20350913
11076861

REFERENCE
AUTHORS
4 (sites)
The RIKEN genome Exploration Research Group Phase II Team and the FANTOM Consortium.

TITLE
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5 (bases 1 to 3063)

JOURNAL
AUTHORS
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hirooka, T., Horii, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurimura, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, K., Sakai, K., Sano, H., Sasaki, D., Schirml, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Teijima, Y., Toyota, T., Yamamura, T., Yamanaka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, url: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGATTCGAGTCAATTAATATCCCCCCCCCC 3'], cDNA was prepared by using trihalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot - 10.0 and subtraction to Rot - 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTCAATTAATATCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised between 3 and 7 kb was selected before cloning. Vector: a modified pluscript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: Salt; 3' end: BamHI. Host: DH10B.

FEATURES
source
Location/Qualifiers
1..3063
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="MGD:MG1:1907565"
/db_xref="taxon:10090"
/clone="493244H21"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
438..1388
/note="data source: SPCR, source key: Q9BPFB, evidence: ISS putative
similar to T1 OLFATORY RECEPTOR"
/codon_start=1
/protein_id="BAB30304.1"
/db_xref="GI:12855358"
/translation="MEPMNSTLESGLTIVGILDSGSPFLLCATVTLTLMALLISNGL"

CDS

Query Match 19.0% Score 175.4; DB 11; Length 3063;
Best Local Similarity 49.6%; Pred. No. 4e-34;
Matches 449; Conservative 0; Mismatches 456; Indels 0; Gaps 0;

polyA_signal
3038..3043
/note="putative"

polyA_site
3063
/note="putative"

BASE COUNT 879 a 612 c 696 g 876 t
ORIGIN

3 gaatcacagcgttgtaactgagtcattatctatctgagcctcaacaaagcctgactcca 62
|||||
449 GAATCTACCTTGGAAGTGGATTCATCTGTGGGGAATCTGGATGGCAGTGCCTTCC 508
63 gggaattatctctctctttttctcatctgctatcttggtgctttctcggcaaatgct 122
|||||
509 TGAAGTCTGCTGTGGCCACAGTTACAACTCTTACATGTTGGCAGCTGATCAGCAATGAGCT 568
123 catcatcattgccaatatactatagcaaaccttgatacagccatgatagtgtttccttct 182
|||||
569 TCTACTCCCGGTCACACAGATGATGCCGCTTACCTACCTACAGTACCTCTCACTAG 628
183 gaaactgctgtgtggaatcatctgcaacaagcattacacgaatgctggtggac 242
|||||
629 GCAGCTGTCTCATGATGACCTCTCTTCACATGTTGATCCACACAGTGTGTGGA 688
243 catgctaacatagaataatcatctatcatatgagcgtgcatgtgtccagctcttctgtt 302
|||||
689 TTCTTCGTCACAGACAAACCATATCTTGAAGGATGCTTCATGTTCTCTCAGC 748
303 cacatgctctcgtggagcgtgagatgcttctcaccacacatgactgacagctatgt 362
|||||
749 AATGACATTTGGGTGGTGGCAGAGAGCTCTTCTGGCTTACGCTATGATGATGTGT 808
363 ggcacattgttctcctctcatcatcatcatatgaaacacacatatgtgtgacct 422
|||||
809 GGCATTTGTCATCTCTTAATCAATGATCTTCAATGATGATGATGATGATGATGAT 868
423 gctcagatgctatgctatctgcatgacacacacacacacacacacacacacacacac 482
|||||
869 GGTGGCATATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 928
483 gaggtgactctctgtgtggccaacacacacacacacacacacacacacacacacac 542
|||||
929 GCACCTTCCCTTTTGGCATGTCAGGAATTCAGACACCTGCTGTGAGGTTCTCTCAT 988
543 gctgcttgctcctgtgagccctgtgaatcatgagtgagtgatgtgtgtgatat 602
|||||
989 GTTGAATTTGGCTTGTCCAGACACATCTCAATATGACCTATGTTATGATGACAGAGCT 1048
603 taacctggccatagggagacttattcttaccatgcatctcctaagtgttcaattgttgc 662
|||||
1049 GATATCTTATTTGCTCCCTCTGTCATTTATTTACCTCTACCTACCTATTTCTATTTAC 1108
663 tatctccgatacgcacagtagaaggaaggaaggaaggaaggaaggaaggaaggaagga 722
|||||
1109 TGTGCTGCATGCTTCAATTAATGAGGGGAGAAAGCCCTTGTACCTGTTCTCTCCCA 1168
723 tctcagatggtgaccttacttctcctgtaatactacacatacgcacacacacacacac 782
|||||
1169 CTTGACTGTGGTGGAGTCTTATGAGGGGAGCCACTTATGATGATGATGATGATGATGAT 1228
783 cagctatacatgaaagagaagaaggtgtgtagctgacactatcatctctgtgactccac 842
|||||
1229 CTTCCACAGTCTTAAGCAAGCAATATCATCTCTGTCTTATGATGATGATGATGATGATGAT 1288

Oy	843	atlaacacggatggctacgcttcagataaggagagctgcaggaagaatgaaggt	902
Oy	1288	TCTGAACCCCTCATTTTACAGCCTGAGAAATGAGAGGTGATTGAGCGCTTTAGAGAGGT	1348
Oy	903	gtttg 907	
		1	
Db	1349	ACTGG 1353	
RESULT	5		
AK017036		1394 bp	linear
LOCUS			HTC 19-JAN-2002
DEFINITION	AK017036	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4933433E02:similar to T1 OLFACTORY RECEPTOR, full insert sequence.	
ACCESSION	AK017036		
VERSION	AK017036.1	GI:12856091	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library clone:4933433E02.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (sites)		
TITLE	Carninci, P. and Hayashizaki, Y.		
JOURNAL	High-efficiency full-length cDNA cloning		
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)		
PUBMED	99279253		
REFERENCE	10349636		
AUTHORS	2 (sites)		
TITLE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, H., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.		
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
MEDLINE	Genome Res. 10 (10), 1617-1630 (2000)		
PUBMED	20499374		
REFERENCE	11042159		
AUTHORS	3 (sites)		
TITLE	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y.		
JOURNAL	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexillary sequencer		
MEDLINE	Genome Res. 10 (11), 1757-1771 (2000)		
PUBMED	20530913		
REFERENCE	11076861		
AUTHORS	4 (sites)		
TITLE	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.		
JOURNAL	Functional annotation of a full-length mouse cDNA collection		
MEDLINE	Nature 409, 685-690 (2001)		
PUBMED	5 (bases 1 to 1394)		
REFERENCE	11076861		
AUTHORS	Atakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Arakawa, T., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schirml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toyota, T., Yamamura, T., Yamana, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.		
JOURNAL	Direct Submission		
PUBMED	Submitted (10-JUL-2000)		
REFERENCE	Yoshihide Hayashizaki, The Institute of		

Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research at Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGGAGAAGAGATCCAGACTCTTTTTCCTTTTTTTTN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGGAGATTCTCGATTAAATTATATCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised between 0.5 and 3 kb was selected before cloning. Vector: a modified pBluescript KS(+) after bulk excision from Lambda PLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.

FEATURES

source

1..1394
location/qualifiers
organism="Mus musculus"
strain="C57BL/6J"
db_xref="MGD:MGI:1894709"
db_xref="taxon:10090"
clone="493343EB02"
sex="male"
tissue_type="testis"
clone_lib="RIKEN full-length enriched mouse cDNA library"
dev_stage="adult"
332..1282
note="data source: SPTK, source key: 09EPF8, evidence: ISS putative similar to T1 OLFACTORY RECEPTOR"
codon_start=1
protein_id="BAB30564.1"
db_xref="GI:12856092"
translation="MEPNNSTLESGFLVINGLDSGSPELLCATVTTLNMLISNGLLLIVTVADRLHPVYILLRQLSLIDLFVSVPNTVDFLLNDNTISEGALOLF SAMLLGGAEDELLAFMYADRYVAICHPNLMIINSPACRILMAVISLALSALGHT VYTHPFPCMSOEIRHLCEVPLPKLACADTSOYELMNVITGVIFLLPSLAITSY SLIIFTVLHMPNSNEGRKKALTVCSSHLTYGMFGAGTFMYVLPSSPHSKQNIISVFYTITPALNDLIYSLEKNKEIVGNRVLRHILPAHAY"

CDS

BASE COUNT 355 a 337 c 304 g 398 t
ORIGIN

Query Match	18.8%	Score 173.8:	DB 11:	Length 1394:
Best Local Similarity	49.5%:	Pred. No. 8.1e-34:		
Matches 448:	Conservative 0:	Mismatches 457:	Indels 0:	Gaps 0:

```

QY      3 gaatacagcgcttgtaactgaattcatatcttcggccacaaagaacctgaaatcca 62
       ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     343 GAATCTCACTTGGAATAATGATTCATTGTGTGGGATTCTGCATGGCATTGCC 402

QY      63 gggaattactcttccttttttcattcattgtctatcttcttggtgccttttcggcaaatgct 122
       ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     403 TGAAGTGTCTGTGGCACAGTTACAACCCTGTATACATGTTGGCAGCTGATACGACAT 462

QY      123 catcatcatctgccaaaatctatagacaaccttgcataagccatgatgtatgttttccttct 182
       ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     463 TCTACTCCTGTGCATCAACAGTGATGCCGGGCTTCACGTACCCATGTACCTCCATCGAG 522

QY      183 gaacatgacctgttgtagacatcatctgacaacaagaatcataccagaagatgctggggagc 242
       ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     523 GCAGCTGCTCTTCATTGACCTTCCTCTTCACATCAGTTGTAACTCCCAACACTGTTTGGA 582

QY      243 catgtctaaccatcagaanaatatcatlcatatagcaggagtgcacgtgccagactctcttgtt 302

```

Db 563 TTTTGTGCTCAGACACACCATATCTCTTGGAGGAGTGGCCCTTCAATGTTCTAGC 642
 Qy 303 cacatgctctcggagctgagatggtctcttcacacacatgctctatgacctatgt 362
 Db 643 AATGACATGGGTGGTGCAGAGACCTCTCTGCGCTTACAGGCTTATGATGATGAT 702
 Qy 363 ggcacattgttcccttcattacactatataaacacacatgctgtagcctt 422
 Db 703 GGCATTTGTCATCTCTTAACATATATCTTCAATGATGATGATGATGATGATGAT 762
 Qy 423 gctcagcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcat 482
 Db 763 GGTGGCATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 822
 Qy 483 gagggttgaattctctgtggcgaacacacacacacacacacacacacacacac 542
 Db 823 GCACTTCCCTTTTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 882
 Qy 543 gctgcttctcctctgtagcctctgtagcctctgtagcctctgtagcctctgtag 602
 Db 883 GTTGAATTTGGCTTGTGCAGACACATCTCAATATGATGATGATGATGATGATGAT 942
 Qy 603 taccctggcacaatgaggaacttattcttcacacacacacacacacacacacac 662
 Db 943 GATATTCCTATTTGCTCCCTCTCTCTGCAATATATATCTCTCTCTCTCTCTCT 1002
 Qy 663 tattctcgtatccgacagtagaagaagaagaagaagaagaagaagaagaagaaga 722
 Db 1003 TGTGCTGACATGCTCTTCAATGAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1062
 Qy 723 tctcagtgtagtacccttactatctctctgtaactctacacacacacacacacac 782
 Db 1063 CTTCAGTGTGGTGGAGATGTTGAGGGGCTGCACTTCTATGATGATGATGATGAT 1122
 Qy 783 cagctatcatttgaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 842
 Db 1123 CTTCACAGTCTTGAAGACATATATATCTCTCTCTCTCTCTCTCTCTCTCTCT 1182
 Qy 843 attaacccagatggtgtagcctcctcagaaatagggagatgagcagcaggaattgga 902
 Db 1183 TCTGACCCCTCATTTACAGCGCTGAGAAATAGAGGATGATGAGGCTGTTAGAGAG 1242
 Qy 903 gtttg 907
 Db 1243 ACTGG 1247

RESULT 6
 AZ255734 726 bp DNA linear GSS 26-JUL-2000
 LOCUS RPCI-23-16619.TJ RPCI-23 Mus musculus genomic clone RPCI-23-16619,
 DEFINITION DNA sequence.
 ACCESSION AZ255734
 VERSION AZ255734.1 GI:9458784
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.
 1 (bases 1 to 726)
 'B', 'Levin', 'M.', 'McGann', 'S.', 'Tsegaye', 'G.', 'Geer', 'K.', 'Krol', 'M.', 'de Jong', 'P.
 and Fraser', 'C.M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 Other_GSSs: RPCI-23-16619.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

FEATURES

source

Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.bufileo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.bufileo.edu/orderframe.htm>)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 plate: 166 row: I column: 9
 Seq primer: SP6
 Class: BAC ends.
 Location/Qualifiers

1..726

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-16619"

/clone_lib="RPCI-23"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site: 1; EcorI; Site: 2; EcorI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcorI and EcorI methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcorI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 167 a 171 c 118 g 270 t
 ORIGIN

Query Match 18.7%; Score 173.2; DB 12; Length 726;
 Best Local Similarity 52.7%; Pred. No. 9.6e-34;
 Matches 376; Conservative 0; Mismatches 338; Indels 0; Gaps 0;

Qy 69 tattctccctttttctcattgctatctgttggccttctcggcaacatgctacat 128
 Db 7 TCTGTTCTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTT 66
 Qy 129 cattgcaaaatctatagacaacacacacacacacacacacacacacacacacac 188
 Db 67 AATTCATTAATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 126
 Qy 189 ggt 248
 Db 127 GTCTTTGATAGACCTCTGTTACTCTGATGTTTACACCCAAATGCTGTAACCTTCA 186
 Qy 249 aacatcagaanaatcattcatatgacagctgcaatgctccagctcttcttcaacatg 308
 Db 187 ATTAATAAGAAATATATATCTTATACGGGGTATGACCAACCTATTTTATTCATT 246
 Qy 309 gttctcggagctgagatggttctctcaccacacacacacacacacacacacacac 368
 Db 247 CTTTGTTATTTCTGAGTGTTATGTTATGTTATGTTATGTTATGTTATGTTATGTT 306
 Qy 369 ttgttccctctcattacagtaactatatagaacacacacacacacacacacacacac 428
 Db 307 CTGTAATCCACTCTTATATATATATGTTATGTTATGTTATGTTATGTTATGTTAT 366
 Qy 429 cagtgatcgtcatatgcaatccacacacacacacacacacacacacacacacacac 488
 Db 367 TGGTTCACTTGTATGTCATTTCTGTTGTCATGTCATGTCATGTCATGTCATGTCAT 426
 Qy 489 gacttctgtggcgaac 548
 Db 427 GACTTCTGTGATGCAACACACATCACTCTCTGATGATCCTCTCTGATGATCCTCT 486
 Qy 549 ttgtcctgttagcctgttaagaatgagtgatgtgtgtgtgtgtgtgtgtgtgtgtgt 608
 Db 487 GCTCTCTGACACGACCTATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 546
 Qy 609 ggcacatggagacatttcttaccatcctcctatgattatgattatgattatgattat 668
 Db 547 TATCCTTGTGCCAGACATCACCATTTTTATCTTATGAGGTTTCACTCTGTCACGACATTT 606

Oy	669	cgcataccgacagatgaagaagcaagaaggccttcacatcatcattccatctaac	728
Db	607	CCACATCACAATCGCATGAGGGCGAAGGTCCACAGGCCTTCACACCTGACACTGATTCCCAATAAT	666
Oy	729	agtgtgatcccttactattctctctgttaattcacacctataaccgcgcgtttc	782
Db	667	TGCTGGTCTCTCTCTTTGCATCAGGTGCATTTAATGAATCTTAACCCCTCCTC	720
RESULT	7		
BH342053			
DEFINITION	BH342053	616 bp	DNA linear GSS 03-DEC-2001
CH230-51L13.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone			
ACCESSION	CH230-51L13		DNA sequence.
VERSION	BH342053		
KEYWORDS	BH342053.1	GI:17272787	
SOURCE	GSS.		
ORGANISM	Norway rat.		
Rattus norvegicus			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
Rattus.			
REFERENCE	1 (bases 1 to 616)		
AUTHORS	Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M.		
TITLE	Rat BAC End Sequences from Library CHORI-230 EcoRI segment		
JOURNAL	Unpublished (1999)		
COMMENT	Other_GSSs: CH230-51L13.TV Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhaoc@tigr.org Clones are derived from the rat BAC library CHORI-230 (http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pjejongsmail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/or erling information.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html plate: 51 row: L column: 13 Seq primer: T7 Class: BAC ends.		
FEATURES			
SOURCE	Location/Qualifiers		
	1..616		
	/organism="Rattus norvegicus"		
	/strain="BN/SSHsd/MCW"		
	/db_xref="taxon:10116"		
	/clone="CH230-51L13"		
	/clone_lib="CHORI-230 Segment 1"		
	/sex="Female"		
	/cell_type="Brain"		
	/note=Vector: pPARBAC2.1; Site_1: EcoRI; site_2: EcoRI; CHORI-230 Rat (BN/SSHsd/MCW) BAC library produced by Pieter de Jong"		
BASE COUNT	142 a 152 c 108 g 214 t		
ORIGIN			
Query Match	18.6%; Score 172; DB 12; Length 616;		
Best Local Similarity	57.0%; Pred. No. 1,8e-33;		
Matches 333; Conservative 0; Mismatches 250; Indels 1; Gaps 1;			
Oy	164	ccatgtatgtttctctcttgacactcgctgtgttggaatcatctgcacaacaagatca	223
Db	1	CAAATGACTCTTTTCCCTTTTAACCTTGATCATTTGTTGACCTCGTAGCTTCACAGTTCCA	60
Oy	224	taccgaagatcgtgggacacatgcataacaaagaataacatcatcatatgcaggtcga	283
Db	61	CACGCCAAATAGCTGATGA-GTTTAAATATGAGAGAGAACATATTATACCTACCAAGAATGCA	119

OY	284	tgccacagctctctgtgttcacaaibgctccgagagcgaagtgctccctccacaca	343
Db	120	tgacacagcttcttcttcttcttccagtcttcttcgcatttctgacgttatgtcgtacgttga	179
OY	344	tgagctatgacccgctatgttgccatttgttccctcttcatacagtaactatagaacc	403
Db	180	tggcctatgattgattgctatgagccactctgtaaccactattgtacaaattgtcagtgctc	239
OY	404	accatactgtgttagccttgctccagcaagtgatagcctatttgagtaacaaattccctgg	463
Db	240	cgaaattgtatggttgaacctgcgttgccctctacataatggcatttcttgctggcattgg	299
OY	464	tgcaacagcctcttcatcaatgaagtgaactcttcgtgggccaacaaacttgaacctct	523
Db	300	ctcacacagggcgatgctgacacgtgactcttctgtgatgacaaacactatcatctact	359
OY	524	tctgtgagatacccccatctgcttgctctgtccctgtagccctgtaagaatcaatgaagtga	583
Db	360	tctctgacatcccccttctgtgcttcaactcttctgcacacgcactatgacatgacgtgg	419
OY	584	tgtgtatgtatgcgatatattaccctgcgcataagggagcttatcttcaocgacatcct	643
Db	420	aggtctttgtgtttgtataggacatcaacatcttctgtgcccacgactacaccttcttctcct	479
OY	644	atggttataccttgcttgctatcttcgtaatccgacagtaagaagcaagaagaagcct	703
Db	480	atggtttatctccctcctcaagatatttttcaattatcagctcccaagaaagcaggtccaggcct	539
OY	704	tctcaacatgctcatctcatctcccaagaagtgagaccttaccat	747
Db	540	tcagcacgtgcagttccacataatccgaagtttctctgtgtttt	583

RESULT	8
LOCUS	BGI93339 762 bp mRNA linear EST 21-APR-2001
DEFINITION	RS112467 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION	BGI93339
VERSION	BGI93339.1 GI:13715026
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 762)
AUTHORS	Harrington,J., Sherf,B., Randlett,S., Jackson,P.D., Perry,R., Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Boozer,S., Mays,R., Smith,E., Veloso,N., Kikla,A., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J. and Ducar,M.
TITLE	Creation of genome-wide protein expression libraries using random activation of gene expression
JOURNAL	Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE	21227151
COMMENT	Contact: Scott J. Cain Athersys, Inc. 3201 Carnegie Ave, Cleveland, OH 44115, USA Tel: 216 431 9900 Fax: 216 361 9596 Email: scaina@atersys.com High quality sequence stop: 551. location/Qualifiers 1. 762 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="Athersys RAGE Library" /cell_line="HT1080" /note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily
FEATURES	
source	

THIS PAGE BLANK (USPTO)


```

1 COUNTRY: US
2 ZIP: 60601-6780
3 COMPUTER READABLE FORM:
4 MEDIUM TYPE: Floppy disk
5 COMPUTER: IBM PC compatible
6 OPERATING SYSTEM: PC-DOS/MS-DOS
7 SOFTWARE: PatentIn Release #1.0, Version #1.25
8 CURRENT APPLICATION DATA:
9 APPLICATION NUMBER: US/08/748,506
10 FILING DATE: 08-NOV-1996
11 CLASSIFICATION: 435
12 PRIOR APPLICATION DATA:
13 APPLICATION NUMBER: US 60/033,751
14 FILING DATE: 09-NOV-1995
15 CLASSIFICATION: 435
16 ATTORNEY/AGENT INFORMATION:
17 REFERENCE/DOCKET NUMBER: 74940
18 TELECOMMUNICATION INFORMATION:
19 TELEPHONE: 312-616-5600
20 TELEFAX: 312-616-5700
21 INFORMATION FOR SEQ ID NO: 5:
22 SEQUENCE CHARACTERISTICS:
23 LENGTH: 966 base pairs
24 TYPE: nucleic acid
25 STRANDEDNESS: double
26 TOPOLOGY: linear
27 MOLECULE TYPE: DNA (genomic)
28 US-08-748-506-5

```

```

seq.documentation.block:
Sequence 5, Application US/08748506
Patent No. 6159707
GENERAL INFORMATION:
APPLICANT: Ronnett et al.
TITLE OF INVENTION: NOVEL SPERM RECEPTORS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leydig, Voit & Mayer, Ltd.
STREET: Two Prudential Plaza, Suite 4900
CITY: Chicago

```

```

2 AsnHisSerValValThrGluPheLeuLeuLeuGluGlyLeuThrLysLysSpr 18
3 .....
31 AAATAGTTTGGACGTGCAACACATT.....GCATTGGCCAAAGTTTTC 71
18 GcLuuGluGngLylLeuLeuPheLeuPhePhe.....LeuLeuValT 32
72 TGAAGTCCCTGGAGAAATGCTTCTCCCTGGTCAACCTCATCTTCATCT 121
32 yLeuValAlaPheLeuGluGlyAsnMetLeuLeuLeuLeuLeuAlaLysLeuTyr 48
122 TCTTAGTATCACTAACAGAAATGCTCATAGTCTGGTATTATTGATACC 171
49 SerAsnThrLeuHisThrProMetLysValPheLeuLeuThrLeuAlaVal 65
172 AGTCATCTCTACACACCCCATGACTTCTTGGGCCAACTTGCTCT 221
65 ValAlaPheLeuGlyThrThrSerLeuLeuLeuLysMetLeuGlyThrM 82
222 CCTGGAGATTGGCTACTATGCTGTGATACCCAAAGTGGCTGACAGGCC 271
82 etLeuThrSerGluAsnThrIleSerTyrAlaGlyCysMetSerGlnLeu 98
272 TTGTGATGACGCCAGAGAGATCTCTGGAGGGATGGCCACACAGATG 321
99 PheLeuPheThrThrIleSerLeuGlyAlaGluMetValLeuPheThrPhe 115
322 TTTTATTTGGCAATTTTGGTATTAAGTGGATGCTGCTCATTTGGCAGCAT 371
115 tAlaTyrAspArgTyrValAlaIleCysPheProLeuHisTyrSerThrT 132
372 GCGCTTTGACCGGTGATGGCATATGCTCCCACTCCATCATCAACACC 421
132 LeuMetAsnHisIleMetCysValAlaIleLeuLeuSerMetValMetalAla 148

```

[illegible]

```

: REFERENCE/DOCKET NUMBER: 74940
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312-616-5600
: TELEFAX: 312-616-5700
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 966 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-748-506-7

alignment_scores:
      Quality: 620.50      Length: 296
      Ratio: 2.846      Gaps: 3
Percent Similarity: 73.649      Percent Identity: 44.257

alignment_block:
US-09-975-308-9 x US-08-748-506-7 ..

Align seg 1/1 to: US-08-748-506-7 from: 1 to: 966

16 lyslysproglnleugllyllellepheleuphe.....le 29
||| ||||| ||| ||||| |||
64 AAGTCTCTGAGGTCCCGGAAATGCTCTCTCCGTACCCCAACCT 113
29 ulleValTyrleuValAlaPheleuGlyAsnMetleullelleAla 46
: : : : : : : : : : : : : : : : : : : : : : : :
114 TCTCATGTTCTTAGTACACTACACAGAAATACCTCATACCTTCTA 165
46 ysleTySerasnThrleuHisThrProMetTyrValPheleuThr 62
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
164 TTGTGACACAGTCACTCTACACACCCCACTGTTCTTGTGGCAAC 213
63 leuAlaValAlaAspIleleCysThrThrSerlleleProlysmet 79
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
214 TTGTCTCTCTGAGATGGCTATATCTGCTGTGTCATACCCAGATGCT 265
79 uGlyThrMetleuThrSerGluAsnThrIleSerTyrAlaGlyCys 96
| : : : : : : : : : : : : : : : : : : : : : :
264 GAGAGACCTTGGAGTGAAGGCCGAGGAGATCTTGGAGGTTGTCCT 313
96 erglnleuPheleuPheThrTrpSerleuGlyAlaGluMetValleuPhe 112
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314 CACAATGTTCTTCTCATATTTCTTGGTATACAGTGTGCTGCTATG 363
113 ThrThrMetAlaTyrAspArgTyrValAlaIleCysPheProleuHis 129
: : : : : : : : : : : : : : : : : : : : : : :
364 GAGCGATAGCCCTTGGACCGCATATATGCTATATGTTCCACATCTCA 413
129 rSerThrIleMetAsnHisHisMetCys.....ValAlaLeuSer 144
| : : : : : : : : : : : : : : : : : : : : : :
414 TGCACACCCGAATGAGTCTGGGGTATGTGCTTATTTGGCAATTTGCAT 463
144 eValMetAlaIleAlaValThrAsnSerTyrValHisThrAlaLeu 166
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
464 GGGTATGGAGCACTATTA.....GGTCTGGAGACAGCAAAATTTAT 507
161 MetArgLeuThrPheCysGlyProAsnThrIleAspHisPhePheCys 177
: : : : : : : : : : : : : : : : : : : : : : :
508 TTCTCTCTGAACTTGTGTGAACCGTGTAGATAGACCACTTCTGTGTA 557
177 ulleProProleuAlaIleuSerCysSerProValArgIleAsnIle 194
: : : : : : : : : : : : : : : : : : : : : : :
558 CCTTCCACCTCTCCGGGACCTTGGCTGTGGATACATCCCAAAATAGG 607
194 alMetValTyrValAlaAspIleThrleuAlaIleGlyAspPheIle 210
: : : : : : : : : : : : : : : : : : : : : : :
608 CTGCATCTTTGTGGACACAGTGTGCTGTGCATATTATGCAATTTTAC 657
211 ThrCysIleSerTyrGlyPheIlelleValAlaIleleuArgIleArg 227

```

```

|||||
658 ATCATTTCTTCATGTCAGAAATTCGCTGTCAGTGTGTCATGTCCTTC 707
227 rValGluglyLysArglySalaphSerThrCysSerSerHisLeuThr 244
: |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
708 ACCGAGGGGGCCGCAATGAAGCTCTCTACCTGTTCATCTCACCTACTG 757
244 aValThrLeuThrTyrrSerProValIleTyrrThrTyrrIleArgProAla 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
758 TAGCACACTCTCTATGCTGCAACATCTGCACCTATTGAGGTCCAG 807
261 SerSerTyrrThrPheGluArgAspLysValAlaIleAlaLeuThrPhe 277
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
808 TCTAGCCACTCACCCAGAGAGTGGACAACCTGCGCTCTCTATACATC 857
277 uValThrProThrLeuAsnPrometValTyrrSerPheGlnAsnArgLum 294
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
858 AGTACATCCATGCTGAAATCCCATCATCTACAGCTTAAGGACACAGGAG 907
294 etGlnAlaGlyIleArgLysValPheAlaPheLeuLys 306
908 TAAAGGCTGCACTGAGAAAGAACTCGGGCTGAGAGAAA 945

```

seq_name: /cgn2.6/ptodata/1/ina/6A.COMB.seq:US-08-748-506-8

```

seq_documentation_block:
; Sequence 8, Application US/08748506
; Patent No. 6159707
; GENERAL INFORMATION:
; APPLICANT: Ronnett et al.
; TITLE OF INVENTION: NOVEL SPERM RECEPTORS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Volt & Mayer, Ltd.
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL
; COUNTRY: US
; ZIP: 60601-6780
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/748,506
; FILING DATE: 08-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,751
; FILING DATE: 09-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; REFERENCE/DOCKET NUMBER: 74940
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-616-5600
; TELEFAX: 312-616-5700
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 966 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-748-506-8

```

```

alignment_scores:
  Quality: 614.50      Length: 296
  Ratio: 2.845        Gaps: 3
Percent Similarity: 72.973  Percent Identity: 43.581
alignment_block:

```

US-09-975-308-9 x US-08-748-506-8 ..
Align seg 1/1 to: US-08-748-506-8 from: 1 to: 966

```

16 LysLysProGluLeuGlnGlyIleIlePheLeuPhePhe.....Le 29
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
64 AAGTCTCTGAGAGTCCCTGGAGAAATGCTTCCTGCTGACCTCATCTCT 113
29 uIleValIleThrLeuValAlaPheLeuGlyYasnMetLeuIleIleAla 46
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
114 TCTCATGTTCTTAGTATCTACAGAGAAATGCTCTCATAGCCCTGGCTA 163
46 ySIIleTyrrSerAsnThrLeuHisThrPrometTyrrValPheLeuThr 62
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
164 TTGTGTCACGATCCATCTCTACACACCCCATGTACTCTTCTTGGCCAC 213
63 LeuAlaValAlaAspIleIleCysThrThrSerIleIleProLysMetLe 79
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
214 TTGTCCTCTCTGGAGATGGCTATACCTTGTCTGCTCTGTATCCACATG 263
79 uGlyThrMetLeuThrSerGluAsnThrIleSerTyrrAlaGlyCysMet 96
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
264 GCAGAGCTCTGTGAGTGGAGGCGGAGAGATCTTACAGTGGGATGGCCA 313
96 eArgLLeuPheLeuPheThrThrTrpSerLeuGlyAlaGluMetValLeuPhe 112
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
314 CACAGATGTTTCTTCTCATATCTTGTGTAATACAGTAGTGTGCTTATG 363
113 ThrThrMetAlaTyrrAspArgTyrrValAlaIleCysPheProLeuHis 129
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
364 GCAGCCATGGCTTGTGACCCCTATATGCTATATGTTCCCACTCCACTA 413
129 rSerThrIleMetAsnHisHisMetCys.....ValAlaLeuLeuSer 144
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
414 TGCACCCGAAATGAGTGTGAGATGTGCTGCCACTTGGCAATGTTTCAT 463
144 eValMetAlaIleAlaValThrAsnSerTrpValHisThrAlaLeuIle 160
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
464 GGTGATGGATGCATAGTA.....GCTGCGACAGACACCAATTTAT 507
161 MetArgLeuThrPheCysGlyProAsnThrIleAspHisPhePheCysG 177
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
508 TTTCCTTGAACTCTGTGACCTGTGAGATACACCACTCTTCTGTGA 557
177 uIleProProLeuLeuAlaLeuSerCysSerProValArgIleAsnGlu 194
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
558 TCTTCCACCTCTCTGCGACTTGTGCTGTGTGATACATCCCAATGTAG 607
194 alMetValTyrrValAlaAspIleThrLeuAlaIleGlyAspPheIleu 210
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
608 CTGCCATCTTGTGTAGTATGCTCTGCTGCAATATAGCCCTTTTGTCTG 657
211 ThrCysIleSerTyrrGlyPheIleIleValAlaIleLeuArgIleArg 227
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
658 ATCATTTATCTTTATGTCAGAAATTCGCTGCAAGGCTGAGTGCCTTC 707
227 rValGluglyLysArglySalaphSerThrCysSerSerHisLeuThr 244
: |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
708 ACCGAGGGGGCCGCAACAGCCCTTCAACCTGTCTCCACCACTACTTG 757
244 aValThrLeuThrTyrrSerProValIleTyrrThrTyrrIleArgProAla 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
758 TAGTCACACTCTTTATGCTGCTCAGATCTGTATCTATTGAGGCTAAG 807
261 SerSerTyrrThrPheGluArgAspLysValAlaIleAlaLeuThrPhe 277
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
808 TCTAGCCACTCACCCAGAGTGGACAACCTTGGCGCTCTCTACACAGC 857
277 uValThrProThrLeuAsnPrometValTyrrSerPheGlnAsnArgLum 294
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
858 AGTACATCCATGTTGAAACCTATCATCTATAGTTAAGGACACAGGATG 907
294 etGlnAlaGlyIleArgLysValPheAlaPheLeuLys 306

```

908 TCAAGCAGCAGCTGAGAGATTCTGGCCCTGAGAAAA 945

seq_name: /cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-827-291A-1

seq_documentation_block:

; Sequence 1, Application US/08827291A

; Patent No. 5874243

; GENERAL INFORMATION:

; APPLICANT: Macina, Roberto

; APPLICANT: Sathe, Ganesh

; TITLE OF INVENTION: NOVEL OLRCCL5 RECEPTOR

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SmithKline Beecham Corporation

; STREET: 709 Swedeland Road

; CITY: King of Prussia

; STATE: PA

; COUNTRY:

; ZIP: 19406

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/827,291A

; FILING DATE: 28-MAR-1997

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: King, William T

; REGISTRATION NUMBER: 30,954

; REFERENCE/DOCKET NUMBER: GP50001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 610-270-5015

; TELEFAX: 610-270-5090

; TELETYPE:

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1290 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; US-08-827-291A-1

; alignment_scores:

; Quality: 598.00

; Ratio: 2.682

; Percent Similarity: 73.841

; Percent Identity: 36.755

; alignment_block:

; US-09-975-308-9 x US-08-827-291A-1 ..

; Align seg 1/1 to: US-08-827-291A-1 from: 1 to: 1290

; 2 AsnHisSerValValThriGluPheLeuLeuLeuGlyLeuThrLysLysSpr 18

; 308 AATTGCACTTCAATCCGACTTCATCTTCTGGAAATCTTCAATGACAG 357

; 18 ogIuLeuGlnGlyLeuLeuPheLeuPheLeuLeuLeuValTyrLeuValA 35

; 358 CCCACCCACACACTTCTCTTCTTCTGATGATGATGATGATGATGATG 407

; 35 IarPheLeuGlyAsnMetLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 51

; 408 CATTCAAGGAGAAATCTGTCTATGTTCTCTCATCTTCACTGACACACAG 457

; 52 LeuHisThrGrometTyrValPheLeuLeuLeuThrLeuLeuValValAsp 67

458 CTCACACACCCCATGATCTCTCTCTGACGACCACTGTCCTCCATGACCT 507

68 ...IleIleCysThrThrSerIleIleProLysMetLeuGlyThrMetL 83

508 CATGCTCATCTGCACACACGTACCCAGATGGCTTCAACATCAC 551

83 eutHisSerGluAsnThrIleSerTyrAlaGlyCysMetSerGlnLeuPhe 99

552 TGCTTGACGAGCAAGTCATTTGATGGCTGTGTGGACACAAATTTTC 601

100 LeuPheThrThrPheSerLeuGlyAlaGlyMetValLeuPheThrMetAl 116

602 TTTATATACATCACCTGCTGCTGATGATGCTTTTGTGCTGTATGCG 651

116 ATYAspArgTyrValAlaIleCysPheProLeuHisTyrSerThrIleM 133

652 TTATGACCGGTACACTGCTATTCACACCCCTGTAAGATACACCAATCTCA 701

133 etAsnHisMetCysValAlaLeuLeuSerMetValMetAlaIleAla 149

702 TGAGCCCTAAATTTGTGACTTATGACTGCTTTCTGATGCTGGGC 751

150 ValThrAsnSerTyrValHisThrAlaLeuIleMetArgLeuThrPheCyl 166

752 TCTACAGATGAGATCATTTATGCTGTAGCCACATTTCTCTCTCTGCTG 801

166 sGlyProAsnThrIleAspHisPheCysGlnIleProProLeuLeuA 183

802 TGGCTCTCGGGAATGACCCCATCTTCTGTGATGATGATGATGATGATG 851

183 IalLeuSerCysSerProValArgIleAsnGluValMetValTyrValAla 199

852 TCTCTCATGCAATGACACATCAATATTGAAAGTTATTTCATTGTC 901

200 AspIleThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyrG1 216

902 TCTATGATTAATGCTGTTTCCCTGTCGAATCATCATGCTGCTCTATGC 951

216 yPheIleIleValAlaIleLeuArgIleArgThrValGlnGlyLysArgL 233

952 TGGATTAATCTGCTGCTCATTCACATGGATGATGAGAGAGGCTGTCGCA 1001

233 ysalPheSerThrCysSerSerHisLeuThrValValThrLeuTyr 249

1002 AAGCTTTCAGACCTGTTCTCTCATGCTGATGCTGATGCTGATGCTGAT 1051

250 SerProValIleTyrThrTyrIleArgProAlaSerSerTyrThrPheG1 266

1052 GGACGAGTTGTTTCATGATCATGACACCCACATCATGATGCTGCCAAC 1101

266 uArgAspLysValValAlaLeuTyrThrLeuValThrProThrLeuA 283

1102 GCAGGACAACTGGTGTGATTTATTCACACATCTGATCCCATCTGCTGA 1151

283 snProMetValTyrSerPheGlnAsnArgGlnMetGlnAlaGlyIleArg 299

1152 ATCCCTCATCTACAGCTCCGCAACAGAAAGATGACACAGACATTCATG 1201

300 LysVal 301

1202 AAGATC 1207

seq_name: /cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-085-371-5

seq_documentation_block:

; Sequence 5, Application US/09085371

; Patent No. 6218358

; GENERAL INFORMATION:

; APPLICANT: Firestein, Stuart

; APPLICANT: Zhao, Haiding

; TITLE OF INVENTION: Functional Expression of, and Assay for, Functional Cellular

; TITLE OF INVENTION: Vivo

```

197 TyTValAlaAspIleThrLeuAlaIleGlyAspPheIleThrCys11 213
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
613 TTTCTCGGCACATTTTATTTCGTGCGGACCGCTCTGTCTACTGGGGC 662
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
213 eSeTyTcIyPhleIleValAlaIleLeuArgIleArgThrValGlu 230
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
663 ATCTCAATGCGCATCACAGGCTCTGTGATGGGACCCCTCAGCTGCTG 712
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
230 IyLysArgIySalAlaPheSerThrCysSerHisLeuThrValValThr 246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
713 GCCGCCATAAAGCCTTTCAACCTGCGCCGCCACCTCATCTTGATGC 762
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
247 LeuTyTySerProValIleTyThrTyThrTyLeArgProAlaSerSerTy 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
763 ATCTTCTATGCGACCATTTTTCATCATCAAGCCAGGCTTAAGGCAATCC 812
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
263 rThPheGluArgAspLysValValAlaAlaLeuTyThrLeuValThrP 280
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
813 AGCTTTTACACACAAAGCTGGCTGCTGTAGTCACTACGCTGCATTGTAC 862
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
280 rOrhLeuAsnProMetValTySerPheGlnInsArg 292
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
863 CGTTGTCAATCCATCATCTACACTGCTTGCGACCAACAA 900

```


MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 116..1003
US-08-467-948A-1

alignment_scores:
Quality: 577.50 Length: 297
Ratio: 2.831 Gaps: 5
Percent Similarity: 68.687 Percent Identity: 42.424

alignment_block:
US-09-975-308-9 x US-08-467-948A-1 ..

Align seg 1/1 to: US-08-467-948A-1 from: 1 to: 1713

```

2 AsnHisSerValAlaThrGluPheIleIleLeuGlyLeuThrIleLysSpr 18
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
137 AATCAGACGATGTCACAGAGTTCTCTCTGATGATTCCTCCGGGCC 186
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18 OGluLeuGlyGlyIleIlePheLeuPhePheLeuIleValIleuVala 35
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
187 AAGGATTCAGATGCTCTCTTGGGCTCTCTCTCCCTCTCTATGTCATCA 236
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
35 IapHeuGlyAsnMetLeuIleIleIleAlaLysIleIleYrSerAsnThr 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
237 CCCGCTGGGGGAATGGAGCATCTCGGGCTCATCTCATGAGCTCCAGA 286
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
52 LeuHisThrProMetIleValIlePheLeuLeuThrLeuAlaValIaSPII 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
287 CTCACACGCCCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 336
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
68 eIleCysThrThrSerIleIleProLysMetLeuGlyThrMetLeuThr 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
337 CGCCTATGCTGCACACAGTGCACCATGCTGATGACCTCTCTCTCTCATC 386
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
85 eArgLysThrIleSerIleAlaGlyGlyMetSerGlnLeuPheLeuPhe 101
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
387 CAGCCAGGCCATCTCTCTGCTGGTGGATGACATGACATGATCTCTCTTT 436
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
102 ThrTrpSerLeuGlyAlaGluMetValLeuPheThrThrMetAlaIle 118
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
437 TTGAGTTTTCACATCTGATGATGCTCTCTGCTGCTGATGCTCTGCA 486
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
118 ParGlyValAlaIleCysPheProLeuHisIleYrSerThrIleMetAsnH 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
487 TCGTAGCTGGCCATCTGCACCCCTCTCGATATTTCATCATCATGACCT 536
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
135 IsthMetCysValAlaLeuLeuSerMetValMetAlaIleAlaValThr 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
537 GGAAGCTCTCATCTCATCTG.....GGCATCACT 565
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
152 AsnSerTrp.....ValIsthAlaLeuIle 160
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
566 ...TCCGGACATGCTGCTCCCTCTGCTATGCTGATGATGACCCCAT 612
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
160 eMetArgLeuThrPheCysGlyProAsnThrIleAspHisPhePheCysG 177
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
613 CCAAGACTCCCTTTGCTGGCTCTGTAATCAACACTCTCTCTCTG 662
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
177 IuIleProProLeuAlaLeuSerCysSerProValArgIleAsnGlu 193
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
663 AAATCTGCTGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 712
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
194 ValMetValIleValAlaAspIleThrLeuAlaIleGlyAspPheIle 210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
713 GTGTCATCTTTGAGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 762
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
210 uIleCysIleSerIleGlyPheIleIleValAlaIleLeuArgIleArgT 227
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
763 GGTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 812
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

227 hrValGlyLysArgLysAlaPheSerThrCysSerSerHisLeuThr 243
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
813 CTGGGAGAGCGCCGAGAAAGCCTTCTCCACCTCTCTCTCTCTCTCTCT 862
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
244 ValValThrIleuYrYrSerProValIleIleYrThrIleArgProAl 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
863 GTAGTGGGACTCTCTTGTGAGCGCCATCTCATGATGATGAGCCCTAA 912
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
260 aSerSerIleThrPheGluArgAspLysValVal...AlaAlaLeuYrT 276
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
913 GTCCGCCATCTGAGAGCAGAGAGCTCTTTCTTATTTTACAGT 962
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
276 hrLeuValThrPro...ThrLeuAsnProMetValIleYrSer 288
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
963 TCTTTCAACCCCATGCTTAAACCCCTGATTTACAAC 1001
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

seq_name: /cgn2_6/plodata/1/lna/6A_COMB.seq:US-08-467-947A-1

seq_documentation_block:

Sequence 1, Application US/08467947A

Patent No. 6090575

GENERAL INFORMATION:

APPLICANT: LI, YI

APPLICANT: CAO, LIANG

APPLICANT: NI, JIAN

APPLICANT: GENTZ, REINER

APPLICANT: BUTT, CAROL J.

APPLICANT: SUTTON III, GRANGER G.

APPLICANT: ROSEN, CRAIG A.

TITLE OF INVENTION: Polynucleotides Encoding Human G-Protein

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.

STREET: 1100 NEW YORK AVE., NW, SUITE 600

CITY: WASHINGTON

STATE: DC

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PATENTIN RELEASE #1.0, VERSION #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/467, 947A

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

Prior APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/04079

FILING DATE: 30-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: STEFFE, ERIC K.

REGISTRATION NUMBER: 36,688

REFERENCE/DOCKET NUMBER: 1488.1140002/EKS/KLM

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-371-2600

TELEFAX: 202-371-2540

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1713 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: both

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 116..1003

US-08-467-947A-1

alignment_scores:

Quality: 577.50

Length: 297

Ratio: 2.831 Gaps: 5
Percent Similarity: 68.687 Percent Identity: 42.424

alignment_block:

US-09-975-308-9 x US-08-467-947A-1 ..

Align seg 1/1 to: US-08-467-947A-1 from: 1 to: 1713

```

2 AsnHisSerValValThrgluPheIleIleuLeuGlyLeuThrylsysPr 18
|||||.....|
137 AATGACAAATGTCACAGAGTTCTCTCTACTGGAGATTCTCTGGGCC 186
|||||.....|
18 oGluLeuGlnGlyIleIlePheLeuPhePheLeuIleValThryLeuVala 35
|||||.....|
187 AAGGATTCAGATGCTCTCTTGGGCTCTCTCTCTCTCTCTCTCTCA 236
|||||.....|
35 lApheLeuGlyAsnMetLeuIleIleIleAlaLysIleTySerAsnThr 51
|||||.....|
237 CCTCTGGGGAATGGACACCTCTGGGCTCATCTCACTGGACCTCAGA 286
|||||.....|
52 LeuHisThrProMetTyValPheLeuLeuThryLeuAlaValAsp11 68
|||||.....|
287 CTCACACACCCATGTAATGCTCTCTCTCTCTCTCTCTCTCTCTCA 336
|||||.....|
68 eIleCysThrThrSerIleIleProLysMetLeuGlyThrMetLeuThrs 85
|||||.....|
337 GCGCTATGCTGCAACACAGTCCCTCATGCTGCTGCACTCTCTCTC 386
|||||.....|
85 eGluAsnThrIleSerTyValIleGlyCysMetSerGlnLeuPheLeu 101
|||||.....|
387 CAGCCAAACCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 436
|||||.....|
102 ThrTrpSerLeuGlyAlaGluMetValLeuPheThrThrMetAlaTy 118
|||||.....|
437 TTGAGTTTGCACATGTAATGCTCTCTCTCTCTCTCTCTCTCTCA 486
|||||.....|
118 PaGlyTyValAlaIleCysPheProLeuHisTySerThrIleMetAsn 135
|||||.....|
487 TCGGACTGCGCATCTGCAACCTCTCTCTCTCTCTCTCTCTCTCA 536
|||||.....|
135 lShMetCysValAlaLeuLeuSerMetValMetAlaIleAlaValThr 151
|||||.....|
537 GGAAGTGTGCATCCTCTG.....GGCATCT 565
|||||.....|
152 AsnSerTrp.....ValHisThrAlaLeu1 160
|||||.....|
566 ..TCCTGGACATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 612
|||||.....|
160 eMetArgLeuThrPheCysGlyProAsnThrIleAspHisPheCysG 177
|||||.....|
613 CTTAGACTGCTCTTGTGGCTCTCTCTCTCTCTCTCTCTCTCTCT 662
|||||.....|
177 lIleProLeuLeuAlaLeuSerCysSerProValArgIleAsnGlu 193
|||||.....|
663 AAATCCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 712
|||||.....|
194 ValMetValTyValAlaAspIleThrLeuAlaIleGlyAspPheIle 210
|||||.....|
713 GTGCTCATCTTTGAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 762
|||||.....|
210 uThrCysIleSerTyArgIlePheIleIleValAlaIleLeuArgIle 227
|||||.....|
763 GGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 812
|||||.....|
227 hrValGlnGlyLysArgLysAlaPheSerThrCysSerHisLeuThr 243
|||||.....|
813 CTGGGAGAGGCGGCAAGAGCTCTCTCTCTCTCTCTCTCTCTCTCT 862
|||||.....|
244 ValValThrLeuTyTySerProValIleTyThrTyIleArgProAl 260
|||||.....|
863 GTACTGGAGACTCTCTTGTGGAGGCGCATCTCTCTCTCTCTCTCT 912
|||||.....|
260 aSerSerTyThrPheGluArgAspLysValVal..AlaAlaLeuTyTr 276
|||||.....|

```

```

|||||.....|
913 GTCCCGCATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 962
|||||.....|
276 hrLeuValThrProThryLeuAsnProMetValTySer 288
|||||.....|
963 TCTTTTCAACCCGATGCTTAACCCCTGATTTTCAAC 1001
|||||.....|

```

seq_name: /cgn2_6/ptodata/1/lna/6A_COMB.seq:US-08-748-506-9

```

seq_documentation_block:
; Sequence 9, Application US/08748506
; Patent No. 6159707
; GENERAL INFORMATION:
; APPLICANT: Ronnett et al.
; TITLE OF INVENTION: NOVEL SPERM RECEPTORS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Volt & Mayer, Ltd.
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL
; COUNTRY: US
; ZIP: 60601-6780
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/748,506
; FILING DATE: 08-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,751
; FILING DATE: 09-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; REFERENCE/DOCKET NUMBER: 74940
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-616-5600
; TELEFAX: 312-616-5700
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-748-506-9

```

alignment_scores:
Quality: 524.00 Length: 305
Ratio: 2.449 Gaps: 0
Percent Similarity: 70.164 Percent Identity: 33.115

alignment_block:
US-09-975-308-9 x US-08-748-506-9 ..

Align seg 1/1 to: US-08-748-506-9 from: 1 to: 984

```

2 AsnHisSerValValThrgluPheIleIleuLeuGlyLeuThrylsysPr 18
|||||.....|
31 AATGACAAATGTCACAGAGTTCTCTCTACTGGAGATTCTCTGGCCGA 80
|||||.....|
18 oGluLeuGlnGlyIleIlePheLeuPhePheLeuIleValThryLeuVala 35
|||||.....|
81 GCACCTGAAGATCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 130
|||||.....|
35 lApheLeuGlyAsnMetLeuIleIleIleAlaLysIleTySerAsnThr 51
|||||.....|
131 CCTCATGGCAACATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
|||||.....|

```

```

52 LeuHsThrProMetTyrValPheLeuLeuThrLeuAlaValAsp11 68
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 CTGACAGAGCCCATGTAATCTTCTCTCAGACACTTCTTTTGTGGAGTG 230
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
68 eileCysThrSerIleIleProLysMetLeuGlyThrMetLeuThrs 85
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
231 TTGTTTAACTACTGTAATCCCAAGCTCCTCACCATTCTGTGTAG 280
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
85 ergLAsnThrIleSerTyrAlaIleCysMetSerGlnLeuPheLeuPhe 101
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
281 GGAGGCAAAAGATTCCTTGGGGTCTCTCTCAGAGCCCTTCGTCTAT 330
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
102 ThrTrpSerLeuGlyAlaIleMetValLeuPheThrMetAlaTyrAs 118
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
331 CTGTGCGGGGGGCAACAGGTTTTCCTTTGGCTGGTTATCCCTGGA 380
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
118 PArgTyrValAlaIleCysPheProLeuHsIleTyrSerThrIleMetAsn 135
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
381 CCGCTTCTGGCCATCTGCAAAACCTCATATTCCAAACCATCATGAGCC 430
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
135 IShISMetCysValAlaLeuLeuSerMetValMetAlaIleAlaValThr 151
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
431 CAAGGATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 480
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
152 AsnSerTrpValHsThrAlaLeuIleMetArgLeuThrPheCysGlyPr 168
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
481 TTCATGGCCACTCCAGTGTGATGCTTCCAGACATTTTACTGTGGGCC 530
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
168 CAsnThrIleAspHsPhePheCysGlyIleProProLeuLeuAlaLeu 185
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
531 AAACATATTATCTCCACTTTTCTGTGATTTGGACCACTGCAATGCT 580
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
185 erCysSerProValArgIleAsnGlnValMetValTyrValAlaAspIle 201
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
581 CCTGTTCAGAAACCAAGGCTATGAGATGCTTTTACCCTTCCTGCTGA 630
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
202 ThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyrGlyPhe11 218
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
631 ATTGTGCTTTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 680
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
218 eileValAlaIleLeuArgIleArgThrValGluGlyLysArgLysAlaP 235
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
681 AGTAGTCACCATAGTAGACTCCCTCAGCCAGGAGGAGACAGAGAGCTT 730
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
235 heSerThrCysSerSerHisLeuThrValValThrLeuTyrTyrSerPro 251
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
731 TTTCACCTGCTGCTCATCTCATCTCTCTCTCTCTCTCTCTCTCTCTCT 780
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
252 ValIleTyrThrTyrIleArgProAlaSerSerTyrThrPheGlnArgAs 268
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
781 TGTCATTTTATATACCTGAAAGCCAAAGCAGAGAGAGAGAGACACCA 830
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
268 pLysValValAlaAlaLeuTyrThrLeuValThrProThrLeuAsnPro 285
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
831 CAGAGAGAGCTGCTCTGTGAACAGTGTGTGACACCCCTCTGAACCTG 880
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
285 etValTyrSerPheGlnAsnArgIleMetGlnAlaGlyIleArgLysVal 301
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
881 TCACCTACACCTGCGCAACAGAGGTCCACAGGCTCTCAGGAGATGCT 930
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
302 PheAlaPheLeuLys 306
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
931 CTGTCCAGGCTTCA 945
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: /cgn2.6/ptodata/1/lna/6A.COMB.seq:US-08-988-876-2
seq_documentation_block:
; Sequence 2, Application US/08988876
; Patent No. 6063596
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.

```

```

; APPLICANT: Yue, Henry
; TITLE OF INVENTION: G PROTEIN COUPLED RECEPTORS ASSOCIATED
; WITH IMMUNE RESPONSE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,876
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0441 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ. ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1828 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSNOT01
; CLONE: 364702
; US-08-988-876-2

alignment_scores:
Quality: 399.50 Length: 317
Ratio: 1.958 Gaps: 12
Percent Similarity: 64.353 Percent Identity: 37.855

alignment_block:
US-09-975-308-9 x US-08-988-876-2 ..

Align seg 1/1 to: US-08-988-876-2 from: 1 to: 1828

2 AsnHsSerValValThrGluPheIleIleLeuGlyLeuThrLysPr 18
|||:|||||:|||||:|||||:|||||:|||||:|||||:
793 AATCTAACAGATGCTCTATATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 842
|||:|||||:|||||:|||||:|||||:|||||:|||||:
18 oGluLeuGlnGlyIleIlePheLeuPhePheLeuIleValTyrLeuValA 35
|||:|||||:|||||:|||||:|||||:|||||:|||||:
843 AGAAGTCAGAGCCAGTCTCTGCTGGGCTGCTCTCTCTCTCTCTCTCTCT 892
|||:|||||:|||||:|||||:|||||:|||||:|||||:
35 IapheLeuGlyAsnMetLeuIleIleIleAlaLysIleTyrSerAsnThr 51
|||:|||||:|||||:|||||:|||||:|||||:|||||:
893 CGGCTGGGGAGACCTGCTCATCATCTGGGCATCAGCCCTGACTCCAC 942
|||:|||||:|||||:|||||:|||||:|||||:|||||:
943 CTCACACACCCCATGTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 992
|||:|||||:|||||:|||||:|||||:|||||:|||||:
68 eileCysThrThrSerIleIleProLysMetLeuGlyThrMetLeuThrs 85
|||:|||||:|||||:|||||:|||||:|||||:|||||:
993 CGGTTTCACTTCCACACAGGCTCCCAAGATGATGTGTGACATCCAGTCTC 1042
|||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

85  ergluAnthrIleSerTyAlaIleGlyCysMetSerGlnLeuPhe 101
      ::::::::::::::::::::
1043 ACACACAGATCTCTATGACAGCGCCGACTGATGATGCTCTCTT 1092
102  ThrTrpSerLeuGly...AlaGlnMet.ValLeuPheThrThrMetIat 117
      ::::::::::::::::::::
1093 GCCATTTTGGAGGCATGAGAGACATCTCTGATGATGATGCT 1142
117  YRPARATgTyrValAlaIleCysPheProLeuHisTySerThrIleMet 133
      ::::::::::::::::::::
1143 ATGACTGGTTGTGACCATCTGTACCCGCTATATCATATTC ACCATCATG 1191
134  AsnHisHisMetCys.....ValAlaLeuLeuSerMetValMetAlaI 148
      ::::::::::::::::::::
1192 AACCGGTGTCTGTGCTTCTGTAGTTTGTGTGTTTCTTTTCTT...CT 1238
148  eAlaValThrAsnSerTrpValHisThrAlaLeuIleMetArgLeuThr 164
      ::::::::::::::::::::
1239 CAGTCTTTTAACCTCCACAGCTGCACAACTTGATGCTTACAGTGACCT 1288
165  .....PheCysGlyProAsnThrIle 171
      ::::::::::::::::::::
1289 GCTTCACAGATGTGGAAATTCCTTAATTTCTGTGACCTTCACACTC 1338
172  AspHisPhePheCysGlnIleProProLeuLeuAlaLeuSerCysSerPr 188
      ::::::::::::::::::::
1339 TCCCATCTTGCACTG.....TGTGACAC 1361
188  oValArgIleAsnGlnValMetValTyValAlaAspIle.....ThrL 203
      ::::::::::::::::::::
1362 CTTCACCAATTAACATATCATATGATTCCTGCTGCATATGTTGGTTTC 1411
203  euAlaIle...GlyAspPheIleLeuThrCysIleSerTyGlyPheIle 218
      ::::::::::::::::::::
1412 TTCCCATCTTCAGGACCTTTCTCTTACTGTA.....AAAT 1449
219  IleValAlaIleLeuArgIleArgThrValAlaGlnGlyLysArgLysAlaPh 235
      ::::::::::::::::::::
1450 CTTTCTCTCATTTCTGAGGTTTCATCATCATGAGTGGAAATATAACC.TT 1498
235  eSerThrCysSerSerHisLeuThrValAlaThrLeuTyTySerProV 252
      ::::::::::::::::::::
1499 CTCACCTGTGGGTCTCACCTGTGATGTT...TGCTGGAGGTAACCTC 1545
252  aIleTyThrTyIleArgProAlaSerSerTyThrPheGlnArgSp 268
      ::::::::::::::::::::
1546 GGTTCCAGATGTGATCTTCCC.....GAGAAAGG 1577
269  LysValAlaIleAlaLeuTyThrLeuValThrProThrLeuAsnProM 285
      ::::::::::::::::::::
1578 TGCAGTGGCTCAGTATGATGACGAGTGTGACCCCATGCTGAAACCCCT 1627
285  eValTySerPheGlnAsnArgGlnMetGlnAlaGlyIleArgLys 300
      ::::::::::::::::::::
1628 TCATGTACAGCCCTGAGAAACAGGATATGAAAGTCTCCGCGCGG 1674
seq_name: /cgn2_6/ptodata/1/fna/6B_COMB.seq:US-09-439-313-526
seq_documentation_block:
; Sequence 526, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yugu
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig

```

```

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 526
; LENGTH: 963
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-439-313-526

alignment_scores:
  Quality: 374.00      Length: 305
  Ratio: 1.861        Gaps: 9
  Percent Similarity: 65.902  Percent Identity: 30.492

alignment_block:
US-09-975-308-9 x US-09-439-313-526 ..
Align seg 1/1 to: US-09-439-313-526 from: 1 to: 963

9  PheIleIleLeuGlyLeuThrLysLysProGlnLeuGlnGlyIleIlePh 25
   ::::::::::::::::::::
31  TTGTGCTTATTGATG.....CCAGATTAGAGAAAGCCCATTT 71
25  eLeuPhe.....PheLeuIleValTyLeuValAlaPheLeuGlyA 39
   ::::::::::::::::::::
72  CTGGGTTGGTCTCCCTCCCTTCATGATGATGATGATGATGATGATGATG 121
39  smetLeuIleIleIleAlaLysIleTySerAsnThrLeuHisThrPro 55
   ::::::::::::::::::::
122  ACTGATGCTGCTCTTCTCTGTAAGAGACGACGACGCTGACGCTCG 171
56  MetTyValPheLeuLeuThrLeuAlaValAlaAspIleIleCysThrTh 72
   ::::::::::::::::::::
172  ATGTACCTCTTCTCTGATGCTTGCACGACCATGTACCTGCGCTATGCC 221
72  rSerIleIleProLysMetLeuGlyThrMetLeuThrSerGlnAsnThrI 89
   ::::::::::::::::::::
222  ATCCACCAATGCCATAGATCCCTTCTGTTGATTCCTCCAGAGA 271
89  leSerTyAlaGlyCysMetSerGlnLeuPheLeuPheThrTrpSerLeu 105
   ::::::::::::::::::::
272  TTAGCTTTGAGGCTGTCTTACCAGATGTC...TTATTCATGCCCTC 318
106  GlyAla...GlnMetValLeuPheThrThrMetAlaTyAspArgTyVa 121
   ::::::::::::::::::::
319  TGACCATTTGAATTCACCATCTGCTGGCCATGAGCTTGAACGTTATGT 368
121  lAlaIleCysPheProLeuHisTySerThrIleMetLysHisHisMetC 138
   ::::::::::::::::::::
369  GGCCATCTGCCACCCATGCGCCATGCTGCAGTGTCAACAT..... 411
138  yValAlaLeuLeuSerMetValMetAlaIleAlaValThrAsnSerTrp 154
   ::::::::::::::::::::
412  ....ACAGTAACAGCCAGATTGGCATGTGCTGTGTCGCGCGAGTCC 456
155  ValHis.....ThrAlaLeuIleMetArgLeuThrPheCysG 167
   ::::::::::::::::::::
457  CTTCTTTTTCCTCCACTGCTGCTGGCATGACAGCGCTGCTGCA 506
167  yProAsnThrIleAspHisPhePheCysGlnIleProProLeuLeuAla 184
   ::::::::::::::::::::
507  CTCCAATGTCTCTGCTGCTCTGATGTTGTCCACAGATGTAATGAACT 556
184  eSerCysSerProValArgIleAsnGlnValMetValTyValAlaAsp 200
   ::::::::::::::::::::
557  TGCCCTATGACAGACACTTGGCCAAATGTGTATATGTTTACTGCC... 603
201  IleThrLeuAlaIleGly...AspPheIleLeuThrCysIleSerTyrol 216

```

```

111 ||| : : : : : ||| : : : : : ||| : : : : :
604 ATTCTGCTGTCANAGGCGGTGAGCTATGTCATCTCTGCTTCTATT 653
216 yphellelleValAlaIleuArgIleArgThrValGluGlyLysArgL 233
654 TCGATATATGAGAAAGGTTCTGCAACTGCTCCAGCTCAGAGGGGCGCA 703
233 ysaIAPheserThrCysSerSerHisIleuThrValAlaThrLeuTyr 249
704 AGGCTTTGGAACCTGTGTGTACACATGTGTGTGTCACCTGCTTCTAT 753
250 SerProAlaIleTyrThrTyrIle.....ArgProAlaSerSerTyr 264
754 GTGCCACTTATTTGGCCCTCTCATGTTGTACACCGCTTTGGAAACAGCTTCA 803
264 rphGluArgAspLysValAlaAlaIleuTyrThrLeuValThrProT 281
804 TCCCATTTGTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 853
281 hrleuAsnProMetValTyrSerPheGlnAsnArgGluMetGlnAlaGly 297
854 TCATTCATCCCATCATCTATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 903
298 IleArgLysValPhe 302
904 GTGCTGGCTATGTTTC 918

```

seq_name: /cgn2.6/ptodata/1/lna/5A.COMB.seq:US-08-465-980-1

seq_documentation_block:

; Sequence 1, Application US/08465980
; Patent No. 5756309

; GENERAL INFORMATION:

; APPLICANT: Soppet, Daniel R.
; APPLICANT: LI, Yi

; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.

; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPRAT70

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
; ADDRESSEE: STUART & OLSTEIN

; STREET: 6 Becker Farm Road
; CITY: Roseland

; STATE: New Jersey

; COUNTRY: USA
; ZIP: 07068

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,980

; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700

; TELEPHONE: 201-994-1744
; TELEPHONE: 201-994-1744

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 1474 base pairs
; TYPE: nucleic acid

; STRANDEDNESS: single
; TOPOLOGY: linear

; MOLECULE TYPE: cDNA
; FEATURE:

; NAME/KEY: CDS
; LOCATION: 274..1233

US-08-465-980-1

alignment_scores:

Quality: 363.00 Length: 282
Ratio: 1.931 Gaps: 7

Percent Similarity: 66.667 Percent Identity: 31.206

alignment_block:
US-09-975-308-9 x US-08-465-980-1 ..

Align seg 1/1 to: US-08-465-980-1 from: 1 to: 1474

```

29 LeuIleValTyrLeuValAlaPheLeuGlyAsnMetLeuIleIleAl 45
||| : : : : : ||| : : : : : ||| : : : : :
364 CTTTCCATGTATGTAGTGCAGATGTGTGAAACGATCGTGGCTTCAT 413
45 alyeIleTyrSerAsnThrIleuHisThrProMetTyrValPheLeuT 62
414 CGTAGAGACGAGACGACGCTCCGATGACTCTTCTCTGCA 463
62 hrleuAlaValAlaAspIleIleCysThrThrSerIleIleProLysMet 78
464 TGCTTGACGACCATGACCTGACCTTATTCACATCCACCATGCTTAATC 513
79 LeuGlyThrMetLeuThrSerGluAsnThrIleSerTyrAlaGlyCysMe 95
514 CTTGCTCCCTTTTGTGTTGATCCCGAGAGATGAGCATTGAGGCTGTCT 563
95 tSerGlnLeuPheLeuPheThrThrSerLeuGlyAla...GluMetValL 111
564 TAACCAATATGTC...TTTATTCATGCCCTCTCGCATTTGAATCCACA 610
111 eupheThrThrMetAlaTyrAspArgTyrValAlaIleCysPheProLeu 127
611 TCCTGCTGCGCATGCGCTTGTGACCGTTATGTGTGCGCATGCGCACCTG 660
128 HisTyrSerThrIleMetAsnHisHisMetCysValAlaLeuLeuSerMe 144
661 CGCCATGCTCGACGTGCTCAACAAAT.....ACAGTAACAGGCCCA 698
144 tValMetAlaIleAlaValThrAsnSerTyrValHis.....T 157
699 GATTGGCATGCTGCTGTGTGCGCGGATGCCCTCTTTTTCCTACATGCG 748
157 hrAlaLeuIleMetArgLeuThrPheCysGlyProAsnThrIleAspHis 173
749 CTGCTGATGATCAAGCGGCTGCTGTGCAATGATGCTGCTGCTGCTG 798
174 PhePheCysGluIleProProLeuAlaLeuSerCysSerProValArg 190
799 TCCATTGTGTCCACGAGATGATGATGATGATGATGATGATGATGATG 848
190 gIleAsnGluValMetValTyrValAlaAspIleThrLeuAlaIleGly 206
849 GCCCAATGTGATATGCTGTACTGCC...ATTCTGCTGTACTGCGG 895
207 ..AspPheIleLeuThrCysIleSerTyrGlyPheIleIleValAlaIle 222
896 TGGACGTAATGATCATCTCTGCTATTTTGTGATAAAGCAAGACGTT 945
223 LeuArgIleArgThrValGluGlyLysArgLysAlaPheSerThrCysSe 239
946 GTGCATGCTGCTTCCAGATGAGCGGCGCAAGCGCTTTTGGAACTGTGT 995
239 rSerHisLeuThrValValThrLeuTyrTyrSerProValIleTyrThr 256
996 GTCCACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1045
256 yIle.....ArgProAlaSerSerTyrThrPheGluArgAspLysVal 270
1046 CAGTTGTACACCGCTTTGAAACAGCCTTCATCCATTTGCTGTGTGTC 1095

```



```

APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPRAL70
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARILLA, BYRNE, BAIN, GILFILLAN, CECCHI,
ADDRESSEE: STUART & OLSTEIN
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: New Jersey
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07093
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/465,980
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ferraro, Gregory D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-446
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1474 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 274..1233
PCT-US95-07093-1

alignment_scores:
Quality: 363.00 Length: 282
Ratio: 1.931 Gaps: 7
Percent Similarity: 66.667 Percent Identity: 31.206

alignment_block:
US-09-975-308-9 x PCT-US95-07093-1 ..
Align seg 1/1 to: PCT-US95-07093-1 from: 1 to: 1474

29 LeuILEValTYrLeuValAlaPheLeuGlyAsnMetLeuILElleIleAl 45
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
364 CTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 413
45 alySIErTYrSerAsnThrLeuHisThrProMetTYrValPheLeuLeuT 62
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
414 CGTAGAGACGACGACGACGACGACGACGACGACGACGACGACGAC 463
62 hIreuAlaValAlaAspIleIleIleCysThrThrSerIleIleProIyMet 78
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 TGCCTTGACGATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 513
79 LeuGlyThrMetLeuThrSerGluAsnThrIleSerTYrAlaGlyCysMe 95
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
514 CTTGCCCTTTTCGTTGATTCGCCGAGATTAAGCATTAAGAGCCCTCT 563
95 tSerGlnLeuPheLeuPheThrTrpSerIleuGlyAla...GluMetVal 111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
564 TACCAGATGTTC...TTTATTCATGCCCTCAGCCATTAAGATCCACA 610

```

```

111 eupheThrThrMetAlaTYrAspArgTYrValAlaIleCysPheProLeu 127
611 TCCTGTCGCGCATGCGCTTGTACCGTTATGTGGCGATGCGCACCGCTG 660
128 HisTYrSerThrIleMetAsnHisIleMetCysValAlaLeuLeuSerMe 144
661 CGCGATGCTGCAGTGCACACAT.....ACAGTAACAGGCCA 698
144 tValMetAlaIleAlaValThrAsnSerTYrValHis.....T 157
699 GATTGGCATGCTGCTGTGTGCGCGGATCCCTTTTTCCTCCACTGC 748
157 hIreAlaIleMetArgLeuThrPheCysGlyProAsnThrIleAspHis 173
749 CTCGTGCTGATTAACGCGCTGCGCTTCTGCGACCTCAATGCTCTCGAC 798
174 PhePheCysGluIleProProLeuAlaLeuSerCysSerProValAr 190
799 TCCTATGTGTGCCACAGATGTAATGAGTTGGCTATGACAGACACTTT 848
190 gIleAsnGluValMetValTYrValAlaAspIleThrLeuAlaIleGly. 206
849 GCCCAATGTGTATATGATGCTTACTGCC...ATTCTGCTGCTCATGGCG 895
207 ..AspPheIleLeuThrCysIleSerTYrGlyPheIleIleValAlaIle 222
896 TGGACGTAATGTTATCTCTCTGCTTATTTTCGATTAATCGAACGTT 945
223 LeuArgIleArgThrValGluGlyLysArgLysAlaPheSerThrCysSe 239
946 CTGCACACTGCTTCCAGTCAAGTCAGACGGGCCAAGCCTTGGAACTGTGT 995
239 rSerHisLeuThrValAlaThrLeuTYrTYrSerProValIleTYrThr 256
996 GTCAACATGTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1045
256 yIle.....ArgProAlaSerSerTYrThrPheGluArgAspVal 270
1046 CAGTTGTACACCGCTTGGAAACAGCCTTCATCCCATTTGCCGTGTGTC 1095
271 ValAlaAlaLeuTYrThrLeuValThrProThrLeuAsnProMetValTY 287
1096 ATGGGTGACATCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1145
287 rSerPheGlnAsnArgGluMetGlnAlaGlyIleArgLysValPhe 302
1146 TGTGTCCAAACCAACAGATCAACACAGGGGTGCTGCTATGTGTC 1191

seq_name: /cgn2_6/plodata/1/lna/5A_COMB.seq:US-08-671-525B-5
seq_documentation_block:
; Sequence 5, Application US/08671525B
; Patent No. 5703220
; GENERAL INFORMATION:
; APPLICANT: Yamada, Tadataka
; APPLICANT: Gantz, Ira
; TITLE OF INVENTION: Genes Encoding Melanocortin Receptors
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: US
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/671,525B
; FILING DATE: June 27, 1996

```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Deann F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-000853DVB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810)641-1600
TELEFAX: (810)641-0270
INFORMATION FOR SEO ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1080 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1080
US-08-671-525B-5

alignment_scores: Length: 325
 Quality: 185.00
 Ratio: 1.156 Gaps: 12
Percent Similarity: 49.231 Percent Identity: 22.154

alignment_block:
US-09-975-308-9 x US-08-671-525B-5 ..

Align seg 1/1 to: US-08-671-525B-5 from: 1 to: 1080

```
17 LysProGluLeuGlnGlyLeuPheLeuPheLeuLeuValTyrLeu 33
   |||||
232 AAGCCGAGAT.....TTCTGTCCTGGGCA 260
   |||||
33 uValAlaPheLeuGlyAsnMetLeuLeuLeuAlaLysIleTyrSerA 50
   |||||
261 CGTAGCTGCTGGAAACATCCGCTTATCCGCGCTGCAGGAGAG 310
   |||||
50 snrPheHisThrProMetTyrValPheLeuLeuThrLeuAlaVal 66
   |||||
311 GCAACCTGCACCTCCCGATGTTCTTCTCTGCACCGCGGCGGCC 360
   |||||
67 AspIleIleCysThrThrSer.....IleIleProly 77
   |||||
361 GACATGCTGTAAGTGTCCATGACCCCTGAGACATCATGATCGCCAT 410
   |||||
77 smetLeuGlyThrMetLeuThrSerGluAsnThrIleSerTyrAlaGly 94
   |||||
411 CGTCCACAGCGACTACCTGACCTTCGAGGAC.....CAGTTTATCCAGC 454
   |||||
94 yMetSerGlnLeuPhe.....LeuPheThrTyrPseSerLeuGlyAlaGlu 108
   |||||
455 ACATGACACACATCTTCGACCTCATGATCGATCTCCCTGCTGCGCTCC 504
   |||||
109 MetValLeuPheThrThrMetAlaTyrAspArgTyrValAlaIleCysP 125
   |||||
505 ATCTGCACACCTCTGCGCATGCGCCGTCGACAGTACGTCACCATCTTTTA 554
   |||||
125 eProLeuHisTyrSerThrIleMetAsnHis..... 136
   |||||
555 CGGCTCCGCTACACACATCATGACGTAAGGAGGCGCTCACCTTGA 604
   |||||
136 ..... 136
605 TCCTGGCATCTGCTGCTGCGGCTGCTGCGCTGCTGCTCATCTGC 654
137 .....MetCysValAlaLeuLeuSerMetVal 145
   |||||
```

```
655 TACTGGAGACAAATGCTATGTTGCTCATGACATGTTCTTCGC 704
145 lMetAlaIleAlaValThrAsnSerTyrPValHisThrAlaLeuIleMetA 162
   |||||
705 CATGATGCTCCCTCATGGGCGACCCCTCTACGTCACATGTTCTCTTTCGC 754
   |||||
162 rGLeuThrPheCysGlyProAsnThrIleAspHisPhe.....PheCys 176
   |||||
755 GGCTG.....CAGTCACAGCGCATGACA 777
177 GluIleProPheLeuAlaLeuSerCysSerProValAlaGlyIleAsnG 193
   |||||
778 GCACGTCACCT.....GCCGACGGGGTGGCCACAGCAA...CACTC 818
193 uValMetValTyrValAlaAspIleThrLeuAlaIleGlyAspPheIle 210
   |||||
819 ATGCATGAAGGGGCGAGTCACATCACCATTCTCTCGGGCGGTTCATCT 868
   |||||
210 eutHrcysIleSerTyrGlyPheIleIleValAlaIleLeuArgIleArg 226
   |||||
869 TCTGCTGGGCGCCCTTCTCTCCACCTGCTGCTCATCTC..... 909
227 ThrValGluGlyLysArgLysAlaPheSerThrCysSerSerHisLeuTh 243
   |||||
910 .....ACCTGCCCAACC..... 921
243 rValValThrLeuTyrTyrSerProValIleTyrThrTyrIleArgProA 260
   |||||
922 .....AACCCCTACGTCATCTGCTACACTGCCCACT 952
260 lAserSerTyrThrPheGluArgAspLysValAlaAlaLeuTyrThr 276
   |||||
953 TCAACACCTTAC.....CTGGTCTCCATC 975
277 LeuValThrProThrLeuAsnProMetValTyrSerPheGlnAsnArgI 293
   |||||
976 ATGTGCAACTCTGTCATCGAACCCATCATCTTACGCTTTCGGAGCCTGGA 1025
293 uMetGlnAlaGlyIleArgLysVal 301
   |||||
1026 ATTGCGCAACACTTTAGGAGATT 1050
```


OM of: US-09-975-308-9 to: GenEmbl: * out_format : pfs
Date: Jun 24, 2002 12:54 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+R2n model -DEV=xlh
-O=/cgn2_1/USFSP.spool/US0975308/runat_24062002_090303_26237/app-query.fasta.1.368
-DB=Genml1 -GPM=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPEXT=4.500
-GAPEXT=0.050 -XGAPEXT=10.000 -XGAPEXT=0.500 -FGAPEXT=6.000
-FGAPEXT=7.000 -YGAPEXT=10.000 -YGAPEXT=0.500 -DELLOP=6.000
-DELEXT=7.000 -STAR=1 -MATRIX=blom62 -TRANS=human40.cdi
-LIST=45 -DOCALLIG=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pis -NORM=ext -HEAPSIZ=500
-MINLEN=0 -MAXLEN=200000000 -USER=US0975308_CGNI_1.9449
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -NARN_TIMEOUT=30
-NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-975-308-9
Query length: 307
Database: GenEmbl: *
Database sequences: 1797656
Search length: 187333701
Search time (sec): 1843.850000

Score list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_hhg:AC091612	- 1575.00	2435.57	2.7e-127	180657	AC091612 Homo sapiens chromo
gb_pat:AX241862	+ 1574.00	2480.56	8.3e-130	921	AX241862 Sequence 610 from Pat
gb_hhg:AC026038	+ 1574.00	2436.33	4.4e-127	138591	AC026038 Homo sapiens chromo
gb_hhg:AL57039	+ 1574.00	2433.50	3.5e-127	190889	AL57039 Homo sapiens chromo
gb_hhg:AC094718	- 1395.00	2152.23	1.6e-111	160483	AC094718 Rattus norvegicus c
gb_hhg:AC096601	- 1390.00	2125.69	4.9e-110	121466	AC096601 Mus musculus clone
gb_ro:AV073800	+ 1354.00	2132.95	1.9e-110	926	AV073800 Mus musculus olfacto
gb_pat:AX350633	+ 1339.00	2110.56	3.4e-109	798	AX350633 Sequence 53 from Pat
gb_pat:AX350675	+ 1443.00	1802.54	4.5e-92	663	AX350675 Sequence 11 from Pat
gb_hhg:AC101272	+ 954.50	1503.47	2.2e-75	915	AC101272 Mus musculus clone
gb_ro:AV073135	+ 954.50	1476.61	6.9e-74	916	AV073135 Mus musculus olfacto
gb_hhg:AC094870	- 935.00	1423.61	3.2e-71	137438	AC094870 Rattus norvegicus c
gb_hhg:AC094700	- 935.00	1423.67	3.2e-71	137438	AC094700 Rattus norvegicus c
gb_hhg:AC096461	+ 906.00	1377.56	2.3e-68	204143	AC096461 Rattus norvegicus c
gb_ro:AV073132	+ 900.00	1415.65	1.7e-70	930	AV073132 Mus musculus olfacto
gb_ro:AV073176	+ 893.00	1404.59	7.1e-70	930	AV073176 Mus musculus olfacto
gb_ro:AV073062	+ 883.00	1388.79	5.4e-69	930	AV073062 Mus musculus olfacto
gb_hhg:AC094493	+ 873.00	1330.36	9.7e-66	116603	AC094493 Rattus norvegicus c
gb_ro:AV073127	+ 871.00	1369.35	6.4e-68	960	AV073127 Mus musculus olfacto
gb_ro:AV073133	+ 869.00	1366.61	9.3e-68	936	AV073133 Mus musculus olfacto
gb_pat:AX207662	+ 861.00	1351.49	6.5e-67	1241	AX207662 Sequence 30 from Pat
gb_hhg:AC011879	- 861.00	1308.58	1.6e-64	160654	AC011879 Homo sapiens clone
gb_pat:AX512324	- 861.00	1308.16	1.7e-64	168473	AX512324 Human DNA sequence
gb_pat:AX241531	+ 860.00	1352.51	3.7e-67	924	AX241531 Rattus norvegicus c
gb_hhg:AC094700	+ 858.50	1302.81	3.3e-64	157438	AC094700 Rattus norvegicus c
gb_ro:AV073126	+ 857.00	1347.63	1.1e-66	939	AV073126 Mus musculus olfacto
gb_ro:AV073125	+ 852.00	1339.76	2.9e-66	936	AV073125 Mus musculus olfacto
gb_ro:AV073128	+ 851.00	1337.15	3.6e-66	939	AV073128 Mus musculus olfacto
gb_ro:AV074052	+ 850.50	1337.46	3.9e-66	928	AV074052 Mus musculus olfacto
gb_ro:AV073124	+ 850.00	1336.60	4.4e-66	936	AV073124 Mus musculus olfacto
gb_hhg:AC094870	+ 849.50	1279.31	6.8e-63	112730	AC094870 Rattus norvegicus c
gb_ro:AV073154	+ 839.00	1319.02	4.2e-65	957	AV073154 Mus musculus olfacto
gb_ro:AV073578	+ 835.50	1313.80	8.1e-65	924	AV073578 Mus musculus olfacto
gb_ro:AV073134	+ 835.50	1313.66	8.3e-65	939	AV073134 Mus musculus olfacto
gb_hhg:AC094493	- 830.50	1263.22	5.3e-62	116603	AC094493 Rattus norvegicus c
gb_ro:AV073073	+ 816.50	1283.78	3.8e-63	924	AV073073 Mus musculus olfacto
gb_ro:AV073577	+ 755.00	1186.54	9.9e-58	933	AV073577 Mus musculus olfacto
gb_hhg:AC108568	+ 750.00	1137.77	5.2e-55	95794	AC108568 Rattus norvegicus c
gb_ro:AV073818	+ 744.00	1168.88	9.6e-57	963	AV073818 Mus musculus olfacto

gb_ro:AV073328 + 740.00 1162.78 2.1e-56 939 AC091612 Homo sapiens chromo
gb_ro:AV073679 + 732.50 1150.85 9.7e-56 948 AC091612 Homo sapiens chromo
gb_ro:AV073680 + 731.50 1149.21 1.2e-55 954 AC091612 Homo sapiens chromo
gb_hhg:AC095959 - 731.50 1101.03 5.8e-53 224606 AC095959 Mus musculus olfacto
gb_ro:AV073249 + 730.00 1146.98 1.6e-55 939 AV073249 Mus musculus olfacto

seq_name: gb_hhg:AC091612

seq_documentation_block:

LOCUS AC091612 180657 bp DNA linear HTG 05-FEB-2002
DEFINITION Homo sapiens chromosome 1 clone RP11-656022, WORKING DRAFT
SEQUENCE 1 unrounded pieces.
AC091612 AL390860
VERSION AC091612.4 GI:18497169
KEYWORDS HTG; HTGS; PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 180657)
AUTHORS Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 180657)
AUTHORS Kaul, R.K., Olson, M.V., Raymond, C., Clendenning, J., Ivey, R.G. and Haugen, E.D.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
COMMENT On Feb 5, 2002 this sequence version replaced g1:15487406.

Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchgs@u.washington.edu
Drafting Center: SC
Project Information
Center project name: chr-1
Center clone name: RP11-656022 (sc0182)

----- Summary Statistics
Sequencing vector: Plasmid; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 180536 bases at least Q40
Consensus quality: 180557 bases at least Q20
Insert size: 194815; 11.0% error; agarose-fp
Insert size: 180657; sum-of-contigs
Quality coverage: 8.4x in Q20 bases; agarose-fp
Quality coverage: 9.0x in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 180657: contig of 180657 bp in length.
Location/Qualifiers
1. 180657
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-656022"
/clone_lib="RPC1 human BAC library 11"
1. 180657
/note="assembly name:Contig19"

misc_feature
BASE COUNT 52363 a 34237 c 35512 g 58545 t
ORIGIN

alignment_scores:

Quality: 1575.00 Length: 307
 Ratio: 5.130 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.674

alignment_block:

US-09-975-308-9 x AC091612/rev ..

Align seg 1/1 to reverse of: AC091612 from: 1 to: 180657

```

1 MetasnHsSerValValThrGluPheIleIleuGlyLeuThrLysLy 17
|||||
155813 ATGATACACAGCGTGTACAGTTCATTTCTGGGCTCCACCAAAA 155764

17 sProGluLeuGlnGlyIleIlePheLeuPhePheLeuIleValTyrLeu 34
|||||
155763 GCCGAACTCCAGGAAATATCTTCCTTTTCTCATGTCTATCTG 155714

34 aAlaPheLeuGlyAsnMetLeuIleIleIleAlaLysIleTyrSerAsn 50
|||||
155713 TGGCTTTCTCGGCACATGCTCATCATTCGCAAAATCTATPACAA 155664

51 ThrLeuHsThrProMetTyrValPheLeuLeuThrLeuAlaValAs 67
|||||
155663 ACCTTGCATAGCCCATATGTTTCTTCTGACACTGGCTGTGTGGA 155614

67 pIleIleCysThrThrSerIleIleProLysMetLeuGlyThrMetLeu 84
|||||
155613 CATCATCTGCACAAACACATCATACCGAAGATGCTGGGACCATGCTAA 155564

84 hrSerGluAsnThrIleSerTyrAlaGlyCysMetSerGluPheLeu 100
|||||
155563 CATCGAAATAACATTTCATATGACAGGCTGATGCCAGCTCTCTTG 155514

101 PheThrTyrSerLeuGlyAlaGluMetValLeuPheThrThrMetAlaTy 117
|||||
155513 TTCACATGCTCTCGGAGCTGAGATGGTCTCTCACACACAGCCCTA 155464

117 rAspArgTyrValAlaIleCysPheProLeuHsThrIleMetAla 134
|||||
155463 TCACGCGATGCGCCATTGTTCCCTTTCATATACAGTACTATTATGA 155414

134 snHsHsMetCysValAlaLeuLeuSerMetValMetAlaIleAlaVal 150
|||||
155413 ACCACCATATGCTGAGCTTGTCTGACATGCTATGCTATGACAGC 155364

151 ThrAsnSerTyrValHsThrAlaLeuIleMetArgLeuThrPheCysG 167
|||||
155363 ACCAATTCCTGGGTGACACACCTCTATCATGAGGTGACTTCTGTGG 155314

167 yProAsnThrIleAspHsPhePheCysGluIleProProLeuLeuAla 184
|||||
155313 GCCAAACACCATTTGACACTTCTCTGTGAGATACCCCATTTGCTGGCT 155264

184 euSerCysSerProValArgIleAsnGluValMetValTyrValAlaAsp 200
|||||
155263 TGTCCTGTGACCTGTAAAGATCAATGAGGTGATGTATGTCTGTGAT 155214

201 IleThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyrGly 217
|||||
155213 ATTAACCTGCGCCATAGGGGCTTATTTTCTTACCTGCATCTCCATGCTTT 155164

217 eIleIleValAlaIleLeuArgIleArgThrValGluGlyLysArgLysA 234
|||||
155163 TATCATGTGTTCATATCTCCGATCCGACACAGTAAAGGACAGAGAAG 155114

234 laPheSerThrCysSerSerHsLeuThrValValThrLeuTyrTyrSer 250
|||||
155113 CCTTCACACATGCTCATCTCATCTCACAGTGGTGACCCCTTACTATTC 155064

251 ProValIleTyrThrTyrIleArgProAlaSerSerTyrThrPheGlu 267
|||||
155063 CCTGTATCTACACATATATCCGCCCTGCTCCAGCTATACATTGAAAG 155014

```

267 gaSpLyValValAlaAlaLeuTyrThrLeuValThrProThrLeuAsn 284

155013 AGACAAGGTGTAGTGTGACTCTATCTCTGTGATCTCCACATTAAAC 154964

284 rOmEValTyrSerPheGlnAsnArgGluMetGlnAlaGlyIleArgLys 300

154963 CGATGTGTACAGCTTCCAGAAATAGGAGATGCAGCAGGAAATTAGGAG 154914

301 ValPheAlaPheLeuLysHs 307

154913 GTGTTGCATTCTGAAACAC 154893

seq_name: gb_pat:AX241862

seq_documentation_block: 921 bp DNA linear PAT 26-SEP-2001

LOCUS AX241862 Sequence 610 from Patent WO0127158.

DEFINITION AX241862

ACCESSION AX241862

VERSION AX241862.1 GI:15798737

KEYWORDS synthetic construct.

SOURCE synthetic construct.

ORGANISM artificial sequence.

REFERENCE 1 (bases 1 to 921)

AUTHORS Bellenson, J., Smith, D., Lancet, D., Glusman, G., Fuchs, T. and Yanai, I.

TITLE Olfactory receptor sequences

JOURNAL Patent: WO 0127158-A 610 19-APR-2001;

DISCLOSURE (US): YEDA RESEARCH AND DEVELOPMENT COMPANY, LTD. (IL)

FEATURES Location/Qualifiers

source 1..921

BASE COUNT 218 a 235 c 183 g 285 t

ORIGIN

alignment_scores:

Quality: 1574.00 Length: 307
 Ratio: 5.127 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.349

alignment_block:

US-09-975-308-9 x AX241862 ..

Align seg 1/1 to: AX241862 from: 1 to: 921

```

1 MetasnHsSerValValThrGluPheIleIleuGlyLeuThrLysLy 17
|||||
1 ATGAATCACAGCGTGTACTGAGTTCATTTCTGGGCTCCACCAAAA 50

17 sProGluLeuGlnGlyIleIlePheLeuPhePheLeuIleValTyrLeu 34
|||||
51 GGCTAACGCCAGGAAATATCTCTCTTTTCTCATTTGCTATCTG 100

34 aAlaPheLeuGlyAsnMetLeuIleIleIleAlaLysIleTyrSerAsn 50
|||||
51 GGCTAACGCCAGGAAATATCTCTCTTTTCTCATTTGCTATCTG 100

101 TGGCTTTCTCGGCAACATGCTCATCATTCATTCGCAAAATCTATTAACA 150

51 ThrLeuHsThrProMetTyrValPheLeuLeuThrLeuAlaValAs 67
|||||
151 ACCTTGCATAGCCCATGATATTTTCTTCTGACACAGGCTGTGTGGA 200

67 pIleIleCysThrThrSerIleIleProLysMetLeuGlyThrMetLeu 84
|||||
201 CATCATCTGCACACACATCATATACCGAAGATGCTGGGACCATGCTAA 250

84 hrSerGluAsnThrIleSerTyrAlaGlyCysMetSerGluPheLeu 100
|||||
251 CATCAAAAATACCATTTATATGACAGGCTGCATGTGCCAGCTCTTCTTG 300

```

101 PheHrTrpSerLeuGlyAlaGluMetValLeuPheHrThrMetAlaTy 117
 301 TTGCATGAGTCTCTGGAGCTGAGATGTTCTCTTCCACACATGCGCTTA 350
 117 rAspArgTyValAlaIleCysPheProLeuHisTySerThrIleMetAla 134
 351 TGACCGCTATGTGGCATTTGTTCCCTCTTCATATACAGTACGTATGGA 400
 134 snHSHSmcTyValAlaLeuLeuSerMetValMetAlaIleAlaVal 150
 401 ACCACATATGTTGTAGTCCCTGCTGATGATGATGATGATGATGATGATG 450
 151 ThrAsnSerTrpValHisThrAlaLeuIleMetArgLeuThrPheCysG 167
 451 ACCAATTCCTGGGTGCACACAGCTTTATATATGATGATGATGATGATGATG 500
 167 yProAsnThrIleAspHisPhePheCysGluIleProProLeuLeuAla 184
 501 GCCAAACACCATGACCATTTCTCTGTGAGATACCCCATGCTGCTGCTT 550
 184 euserCysSerProValArgIleAsnGluValMetValTyValAlaAsp 200
 551 TGTCCTGATCCCTGTAAGATCATGATGATGATGATGATGATGATGATGAT 600
 201 IleThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyArgI 217
 601 ATTACCTGGCCATAGGGGACTTATTTCTTACTGATCCTCTATGTTT 650
 217 eIleIleValAlaIleLeuArgIleArgThrValGluGlyLysArgLysA 234
 651 TATCATGTTGCTATCTCCGTATCCGACACAGTAAAGCAAGAGAGAGAG 700
 234 lAheserTrpCysSerSerHisLeuThrValValThrLeuTyTrpSer 250
 701 CCTTCATACATGCTCATCTCATCTCATCTCATCTCATCTCATCTCATCT 750
 251 ProValIleTyThrTyTrpIleArgProAlaSerSerTyThrPheGluAr 267
 751 CCTGTATCTACACATATATCCGCTCTTCCAGCATATCATTTGGAAG 800
 267 gAspLysValValAlaAlaLeuTyThrLeuValThrProThrLeuAsn 284
 801 AGCAAGAGGTGTAGTGCACATCTACTCTTGACTCCACATTTAAACC 850
 284 rOmetValTySerPheGlnAsnArgGluMetGlnAlaGlyIleArgLys 300
 851 CGATGCTGATACCTTCCAGATATGAGATGATGATGATGATGATGATGATG 900
 301 ValPheAlaPheLeuLysHis 307
 901 GTGTTGATTTCTGAAACAC 921
 seq_name: gb_htg:AC026038
 seq_documentation_block:
 LOCUS AC026038 138591 bp DNA linear HTG 01-SEP-2000
 DEFINITION Homo sapiens chromosome 1 clone RP11-109C14, WORKING DRAFT
 SEQUENCE 22 unordered pieces.
 ACCESSION AC026038.4 GI:9958133
 VERSION AC026038
 KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 138591)
 TITLE Waterston,R.H.
 JOURNAL The sequence of Homo sapiens clone
 REFERENCE 2 (bases 1 to 138591)
 JOURNAL Unpublished
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (19-MAR-2000) Genome Sequencing Center, Washington

COMMENT

University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 On Sep 1, 2000 this sequence version replaced g1:8567954.
 ----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 ----- Project Information -----
 Center project name: H.NH0109C14
 Summary Statistics -----
 Sequencing vector: p3: 100%
 Chemistry: Dye-primer ET; 100% of reads
 Chemistry: Dye-terminator Big Dye; 0% of reads
 Assembly program: Phrap; Version 0.990319
 Consensus quality: 129539 bases at least Q40
 Consensus quality: 131919 bases at least Q30
 Consensus quality: 133309 bases at least Q20
 Insert size: 140000; agarose-fp
 Insert size: 136491; sum-of-contigs
 Quality coverage: 3.88 in Q20 bases; agarose-fp
 Quality coverage: 4.04 in Q20 bases; sum-of-contigs
 ----- NOTE: This is a 'working draft' sequence. It currently
 * consists of 22 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1 5221: contig of 5221 bp in length
 * 5222 5321: gap of unknown length
 * 5322 9571: contig of 4250 bp in length
 * 9572 9671: gap of unknown length
 * 9672 13673: contig of 4002 bp in length
 * 13674 13773: gap of unknown length
 * 13774 18721: contig of 4948 bp in length
 * 18722 18821: gap of unknown length
 * 18822 24023: contig of 5202 bp in length
 * 24024 24123: gap of unknown length
 * 24124 28988: contig of 4865 bp in length
 * 28989 29088: gap of unknown length
 * 29089 34171: contig of 5083 bp in length
 * 34172 34271: gap of unknown length
 * 34272 40989: contig of 6718 bp in length
 * 40990 41089: gap of unknown length
 * 41090 48355: contig of 7266 bp in length
 * 48356 48455: gap of unknown length
 * 48456 59366: contig of 10911 bp in length
 * 59367 59466: gap of unknown length
 * 59467 61313: contig of 1847 bp in length
 * 61314 61413: gap of unknown length
 * 61414 71241: contig of 9828 bp in length
 * 71242 71341: gap of unknown length
 * 71342 80867: contig of 9526 bp in length
 * 80868 80967: gap of unknown length
 * 80968 96688: contig of 18721 bp in length
 * 96689 96788: gap of unknown length
 * 96789 122864: contig of 23076 bp in length
 * 122865 122964: gap of unknown length
 * 122965 124804: contig of 1840 bp in length
 * 124805 124904: gap of unknown length
 * 124905 126520: contig of 1716 bp in length
 * 126521 126720: gap of unknown length
 * 126721 128799: contig of 2079 bp in length
 * 128800 128899: gap of unknown length
 * 128900 130561: contig of 1662 bp in length
 * 130562 130661: gap of unknown length
 * 130662 132547: contig of 1886 bp in length
 * 132548 132647: gap of unknown length
 * 132648 135428: contig of 2781 bp in length

* 135429 135528: gap of unknown length
* 135329 138591: contig of 3063 bp in length.
Location/Qualifiers

source
1. 138591
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-109C14"

misc_feature
1. 5221
/note="assembly_name:Contig10
clone_end:77
vector_side:left"

misc_feature
5322..9571
/note="assembly_name:Contig11"

misc_feature
9672..13673
/note="assembly_name:Contig12"

misc_feature
13774..18721
/note="assembly_name:Contig13"

misc_feature
18822..24023
/note="assembly_name:Contig14"

misc_feature
24124..28988
/note="assembly_name:Contig15"

misc_feature
29089..34171
/note="assembly_name:Contig16"

misc_feature
34272..40989
/note="assembly_name:Contig17"

misc_feature
41090..48355
/note="assembly_name:Contig18"

misc_feature
48456..59366
/note="assembly_name:Contig19"

misc_feature
59467..61313
/note="assembly_name:Contig20"

misc_feature
61414..71241
/note="assembly_name:Contig21"

misc_feature
71342..80867
/note="assembly_name:Contig22"

misc_feature
80968..99688
/note="assembly_name:Contig23"

misc_feature
99789..122864
/note="assembly_name:Contig23
clone_end:SP6
vector_side:left"

misc_feature
122965..124804
/note="assembly_name:Contig3"

misc_feature
124905..126620
/note="assembly_name:Contig4"

misc_feature
126721..128799
/note="assembly_name:Contig5"

misc_feature
128900..130561
/note="assembly_name:Contig6"

misc_feature
130662..132547
/note="assembly_name:Contig7"

misc_feature
132648..135428
/note="assembly_name:Contig8"

misc_feature
135529..138591
/note="assembly_name:Contig9"

BASE COUNT 44211 a 25199 c 25318 g 41754 t 2109 others
ORIGIN

alignment_scores:
Quality: 1574.00 Length: 307
Ratio: 5.127 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 99.349

alignment_block:
US-09-975-308-9 x AC026038 ..

Align seg 1/1 to: AC026038 from: 1 to: 138591

1 MetcnshtsservAlvalThnGluPhelIleleuGIyleuThrylsy 17
|||||
86786 ATGAATCAGACGGTTGTAAGTCAITATTCTGGGCTCACCAAAA 86835

17 sProGluLeuGInGlyIlellePheLeuPhePheLeuIleValTyLeuY 34
|||||
86836 GCGTCAATCCAGGGAATATCTCTCTTTCTTCTATCTCATCTG 86885
34 aLaIaPheLeuGIyAsnMetLeuIlelleIleAlaIysIleTySerAsn 50
|||||
86886 TGGCTTTCTCGGCACATGCTCATCATCATGCGCAAAATCATTAACAC 86935
51 ThrLeuHstHrPrometyValPheLeuLeuThrLeuAlaValAs 67
|||||
86936 ACCCTGCATACGCCCATGATGTTTCTTCTGACACCTGCTGTGGA 86985
67 pIleIleCysThrThrSerIlelleProIysMetLeuGIyThrMetLeu 84
|||||
86986 CATCATCTGCACACAGCATCATACCGAAGATGCTGGGACCATGCTAA 87035
84 hTserGIuAsnThrIleSeryTyAlaGIyCysMetSergIleuPheLeu 100
|||||
87036 CATCAGAAATACCATTTCTATGACAGGCTGCATGCCAGCTTCTTG 87085
101 PheThrTrpSerLeuGIyAlaIuMetValIleuPheThrMetAlaTy 117
|||||
87086 TTCACATGGTCTCTGGAGCTGAGATGATGTTCTTCCACACATGGCTA 87135
117 rAspArgTyValAlaIleCysPheProLeuHstTySerThrIleMet 134
|||||
87136 TGACCGCTATGTCGCTATGTTCCCTCTCATTAACATCTGTTATGA 87185
134 snHstHsMetCysValAlaIleuLeuSerMetValMetAlaIleAlaVal 150
|||||
87186 ACCACCATATGTTGTAGCTTCTGTCACATGCTCATGCTATGTCAGTC 87235
151 ThAsnSerTrpValHstHrAlaIleuIleMetArgLeuThrPheCysGI 167
|||||
87236 ACCAATTCCTGGGTGACACACAGCTTATCATGATGATGCTTCTGTGG 87285
167 yProAsnThrIleAspHstPheCysGIuIleProLeuLeuAlaIle 184
|||||
87286 GCCAAACACCATGACCCATCTCTGTGATGATACCCCATCTCTGGCTT 87335
184 euserCysSerProValArgIleAsnGIuValMetValTyValAlaAsp 200
|||||
87336 TGCTCTGTAGCCCTGTAAGATCATGAGTATGATGTTGCTGAT 87385
201 IleThrLeuAlaIleGIyAspPheIleLeuThrCysIleSeryGIyPh 217
|||||
87386 ATTAACCTGGCCATAGGAGGACTTATCTTACCTGCATCTCTATGCTTT 87435
217 eIleIleValAlaIleLeuArgIleArgThrValGIyGIyLysArgLysA 234
|||||
87436 TATCATTTGTTCTATTCCTATCCGACAGTAGAAGGCAAGGAAG 87485
234 lApeSerThrCysSerSerHstLeuThrValAlaIleThrLeuTyTySer 250
|||||
87486 CCTTCTCAACATGCTCATCTCATCTCACAGTGGACCTTATCTATCT 87535
251 ProValIleTyThrTyIleArgProAlaSerSeryThrPheGIuArg 267
|||||
87536 CCGTAATCTACCTATATCCGCCCTCTCTCCAGTATATCTGAAG 87585
267 gAspLysValAlaAlaIleuTyThrLeuValThrProThrLeuAsn 284
|||||
87586 AGACAAGGTGTAAGCTGACCTATATCTTGTGCTCCACATTAAC 87635
284 roMetValTySerPheGlnAsnArgGIuMetGlnAlaGIyIleArgLys 300
|||||
87636 CGATGCTGTAAGCTTCCAGATTAAGGAGATGACGACGAGATTTGGAAG 87685
301 ValPheAlaPheLeuTyHst 307
|||||
87686 GTGTTTGCATTTCTGAACAC 87706

seq_name: gb_htg:AL57039

seq_documentation_block: 190889 bp DNA linear HTG 19-SEP-2001

LOCUS AL57039

DEFINITION Homo sapiens chromosome 1 clone RP11-634B7, *** SEQUENCING IN PROGRESS ***, 18 unordered pieces.

ACCESSION AL57039

VERSION AL57039.5 GI:9797648

KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 190889)

AUTHORS Plumb,B.

TITLE Direct Submission

JOURNAL Submitted (18-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

COMMENTS Requests: clonerequests@sanger.ac.uk

On Aug 12, 2000 this sequence version replaced gi:9214044.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: ba634B7

----- Summary Statistics

Assembly program: XGAP; version 4.5

Sequencing vector: plasmid; 108752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 183408 bases at least Q40

Consensus quality: 186557 bases at least Q20

Consensus quality: 188087 bases at least Q20

Insert size: 189189; sum-of-contigs

Insert size: 188721; 6.6% error; agarose-fp

Quality coverage: 4.17x in Q20 bases; sum-of-contigs Quality coverage: 4.30x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently consists of 18 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 12042: contig of 12042 bp in length

* 12043 12142: gap of 100 bp

* 12143 30262: contig of 18120 bp in length

* 30263 30362: gap of 100 bp

* 30363 41423: contig of 11063 bp in length

* 41426 41525: gap of 100 bp

* 41526 48351: contig of 6826 bp in length

* 48352 48451: gap of 100 bp

* 48452 57808: contig of 9357 bp in length

* 57809 57908: gap of 100 bp

* 57909 68643: contig of 10735 bp in length

* 68644 68743: gap of 100 bp

* 68744 75536: contig of 6793 bp in length

* 75537 75636: gap of 100 bp

* 75637 84190: contig of 8554 bp in length

* 84191 84290: gap of 100 bp

* 84291 88033: contig of 4749 bp in length

* 89040 89139: gap of 100 bp

* 89140 94516: contig of 5377 bp in length

* 94517 94616: gap of 100 bp

* 94617 105920: contig of 11304 bp in length

* 105921 106020: gap of 100 bp

* 106021 133564: contig of 27544 bp in length

* 133565 133664: gap of 100 bp

* 133665 137940: contig of 4276 bp in length

* 137941 138040: gap of 100 bp

FEATURES

source

1..190889

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="1"

/clone="RP11-634B7"

/clone_1fb="RP11-11.3"

1..12042

/note="assembly_fragment:01421"

fragment_chain:1

clone_end:77

vector_side:left

12143..30262

/note="assembly_fragment:00925"

fragment_chain:1

30363..41425

/note="assembly_fragment:00451"

fragment_chain:2

41526..48351

/note="assembly_fragment:00297"

fragment_chain:2

48452..57808

/note="assembly_fragment:00082"

fragment_chain:2

57909..68643

/note="assembly_fragment:01434"

fragment_chain:2

68744..75536

/note="assembly_fragment:01941"

fragment_chain:2

75637..84190

/note="assembly_fragment:00841"

fragment_chain:2

84291..89039

/note="assembly_fragment:00517"

fragment_chain:3

89140..94516

/note="assembly_fragment:00194"

fragment_chain:3

94617..105920

/note="assembly_fragment:00685"

fragment_chain:4

106021..133564

/note="assembly_fragment:00244"

fragment_chain:4

133665..137940

/note="assembly_fragment:00993"

fragment_chain:5

138041..141386

/note="assembly_fragment:00049"

fragment_chain:5

141487..143513

/note="assembly_fragment:01137"

143614..159175

/note="assembly_fragment:01189"

159276..174409

/note="assembly_fragment:01199"

174510..190889

/note="assembly_fragment:01519"

BASE COUNT 60637 a 34972 c 35643 g 57919 t 1712 others

ORIGIN

alignment_scores:

COMMENT

Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15624554.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GRP

Center clone name: CH230-5M10

----- Summary Statistics

Assembly program: Phrap; version 0.990329first call to findPhraplist

Consensus quality: 127493 bases at least Q40

Consensus quality: 137458 bases at least Q30

Consensus quality: 144882 bases at least Q20

Estimated insert size: 129461; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently

* consists of 65 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1      5754: contig of 5754 bp in length
*      5755      5854: gap of unknown length
*      5855      11490: contig of 5636 bp in length
*      11491      11590: gap of unknown length
*      11591      17216: contig of 5626 bp in length
*      17217      17316: gap of unknown length
*      17317      21754: contig of 4438 bp in length
*      21755      21854: gap of unknown length
*      21855      25351: contig of 3497 bp in length
*      25352      25451: gap of unknown length
*      25451      30651: contig of 5200 bp in length
*      30652      30751: gap of unknown length
*      30752      34861: contig of 4110 bp in length
*      34862      34961: gap of unknown length
*      34962      38499: contig of 3538 bp in length
*      38499      38500      38599: gap of unknown length
*      38600      41784: contig of 3185 bp in length
*      41785      41884: gap of unknown length
*      41885      45873: contig of 3989 bp in length
*      45874      45973: gap of unknown length
*      45974      50324: contig of 4251 bp in length
*      50325      50324: gap of unknown length
*      50325      53266: contig of 2942 bp in length
*      53267      53366: gap of unknown length
*      53367      56430: contig of 3064 bp in length
*      56431      56530: gap of unknown length
*      56531      59835: contig of 3305 bp in length
*      59836      59936      62863: gap of unknown length
*      59936      62863: contig of 2928 bp in length
*      62864      62963: gap of unknown length
*      62964      65355: contig of 2392 bp in length
*      65356      65455: gap of unknown length
*      65456      68305: contig of 2850 bp in length
*      68306      68405: gap of unknown length
*      68406      70717: contig of 2312 bp in length
*      70718      70817: gap of unknown length
*      70818      73406: contig of 2489 bp in length
*      73407      73406: gap of unknown length
*      73407      77576: contig of 4170 bp in length
*      77577      77576: gap of unknown length
*      77577      79917: contig of 2241 bp in length
*      79918      80017: gap of unknown length
*      80018      83098: contig of 3081 bp in length

```

```

*      83099      83198: gap of unknown length
*      83199      85948: contig of 2750 bp in length
*      85949      86048: gap of unknown length
*      86049      86300: contig of 2252 bp in length
*      86301      88400: gap of unknown length
*      88401      90834: contig of 2434 bp in length
*      90835      90934: gap of unknown length
*      90935      92486: contig of 1552 bp in length
*      92487      92586: gap of unknown length
*      92587      94928: contig of 2342 bp in length
*      94929      95028: gap of unknown length
*      95029      97296: contig of 2268 bp in length
*      97297      97396: gap of unknown length
*      97397      99357: contig of 1961 bp in length
*      99358      99457: gap of unknown length
*      99458      101633: contig of 2176 bp in length
*      101634      101733: gap of unknown length
*      101734      103947: contig of 2214 bp in length
*      103948      104047: gap of unknown length
*      104049      106350: contig of 2303 bp in length
*      106351      106450: gap of unknown length
*      106451      108776: contig of 2326 bp in length
*      108777      108876: gap of unknown length
*      108877      111586: contig of 2710 bp in length
*      111587      111686: gap of unknown length
*      111687      113940: contig of 2254 bp in length
*      113941      114040: gap of unknown length
*      114041      115113: contig of 1073 bp in length
*      115114      115213: gap of unknown length
*      115214      116544: contig of 1241 bp in length
*      116545      116554: gap of unknown length
*      116555      118145: contig of 1591 bp in length
*      118146      118245: gap of unknown length
*      118246      119842: contig of 1597 bp in length
*      119843      119942: gap of unknown length
*      119943      121213: contig of 1271 bp in length
*      121214      121313: gap of unknown length
*      121314      123957: contig of 2644 bp in length
*      123958      124057: gap of unknown length
*      124058      125239: contig of 1162 bp in length
*      125240      125339: gap of unknown length
*      125340      126844: contig of 1505 bp in length
*      126845      126944: gap of unknown length
*      126945      128704: contig of 1760 bp in length
*      128705      128805      128804: gap of unknown length
*      128805      130980: contig of 2176 bp in length
*      130981      131080: gap of unknown length
*      131081      132857: contig of 1777 bp in length
*      132858      132957: gap of unknown length
*      132959      134901: contig of 1944 bp in length
*      134902      135001: gap of unknown length
*      135002      136075: contig of 1074 bp in length
*      136076      136175: gap of unknown length
*      136176      137403: contig of 1228 bp in length
*      137404      137503: gap of unknown length
*      137504      139120: contig of 1617 bp in length
*      139121      139220: gap of unknown length
*      139221      141101: contig of 1861 bp in length
*      141102      141201: gap of unknown length
*      141202      143536: contig of 2335 bp in length
*      143537      144839: gap of unknown length
*      144839      144839: contig of 1203 bp in length
*      144840      144939: gap of unknown length
*      144940      146098: contig of 1159 bp in length
*      146099      146198: gap of unknown length
*      146199      147377: contig of 1179 bp in length
*      147378      147477: gap of unknown length
*      147478      148718: contig of 1241 bp in length
*      148719      148818: gap of unknown length

```

alignment_scores: Quality: 1395.00 Length: 306
Ratio: 4.697 Gaps: 0

Percent Similarity: 97.059 Percent Identity: 85.621

alignment block:
US-09-975-308-9 x AC094718/rev ..

Align seg 1/1 to reverse of: AC094718 from: 1 to: 160483

```

1 Metasnhiservalvalthrgulphellelleleuglyleuhtyrsly 17
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
13579 ATGAACTTCAGAGTGTAAATGAGTTATGATTTGGAGCTTACTCAAAA 13530
17 sprogileugnglyllellepheuelphelellelevaltyrleuv 34
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
13529 ATCGAACTTCAGAGTGTAAATGAGTTATGATTTGGAGCTTACTCAAAA 13480
34 alalaphleuglyasmelueillellelelealyslletyrserasn 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
13479 TAGCTTTCTGTAATTTGTTAAATGTTGTCGCAATTAATTAACACC 13430
51 Thrlleuhtshtrpmettyrvalpheleuelphelellelealvalas 67
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
13429 ACCTTGACACACCACTATATCTCTCTGCGCTGCGTGTGGA 13380
67 pillelecsthrthsrlelleleprolysmelueuglythrmleut 84
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
13379 CATATCTGTACTACACCATCATCCCAAAATGTTGGGAACAAATGTTAA 13330
84 hrsergluasnthrlelethyralaglycysmelsergluelpheleu 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
13329 CTTGCAAAAATTCGATTCATATGAGGGCTGATGCCACCTCTCTTT 13280
101 pthertrpserleuglyalalumevalleuphetrrhmetalaty 117
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
13279 TTTACATGCTCTTAGGGCTGAGATGAGCTTCTTACTACATGCGCTA 13230
117 rasphargtyrvalalalecyspheproleuhtshtrlelemeta 134
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
13229 TGACCCCTATGCGCATTTGCTCCACTGCGTATGACTATTATGA 13180
134 snhishismecysvalalaleuelphelellelelealval 150
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
13179 ACCAATATCATGCTGAGCTTCTGAGATGCTGATGCTATGACAGTA 13130
151 thrasertrpvalalshtraleuilemetargyleuthrphcysgl 167
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
13129 ACCAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 13080
167 typroasnthrleasphishphephecysglulieproleuelpheal 184
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
13079 GCCAAATACCGTGTATCATCTTCTGTGAATATCCCACTGCTGCTT 13030
184 eusercysserprovalargileasngluvalmetvaltyrvalalasp 200
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
13029 TGCTCTGAGTCTGCTGAGAGTATGAGTATGAGTATGAGTATGAGT 12980
201 llerthleuallaleglyaspheilleleuthrthcyslleserlyrgh 217
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
12979 ATGCACTGCTGCTGAGAGTATGAGTATGAGTATGAGTATGAGTAT 12930
217 eillelevalalaleleuargyleargthvalgluglylsarglysa 234
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
12929 TATATGCTGCTATCTCTGCACTCCCTACACGAGAGCAAGAGAAAG 12880
234 lapheserthcysserserhisleuthrvalalathrleuhtyrser 250
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
12879 CCTTCTGACATGCTCATCTCACTCATGCGGTGCTCTTACTATCTCT 12830
251 provallelythrtyrleargproalaserseerlyrthrpheguar 267
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
12829 CCTGTAATCTACACCTATCTGACCTGCACTCCAGCTATACCTTTGAGA 12780
267 gaspysvalalaleuhtyrthleuvalthrprothreusnp 284
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

12779 AGACAGTAGTAGCTGCATATATCTTAGTACTACCTCATATGATC 12730
284 rometvaltyrserphecgluasnarglumetglinalaglylearglys 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
12729 CATAGTATACAGCTTTAGAAAACAAGAGATGACAGTATGAGGAAA 12680
301 valphealaphelenuis 306
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
12679 GATTGTGCACTTTCTTAAA 12662
seq_name: gb_hhg.AC099601

```

```

seq_documentation_block:
AC099601 221466 bp DNA linear HTG 16-NOV-2001
LOCUS AC099601
DEFINITION Mus musculus clone RP23-240K1, WORKING DRAFT SEQUENCE, 3 unordered
pieces.
AC099601.1 GI:16946330
VERSION AC099601.1
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 221466)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-240K1
Unpublished
2 (bases 1 to 221466)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Bana,N., Bastien,V., Boguslavsky,L., Boucknight,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Chapel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Illsey,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kelis,C., Larocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J.,
Menus,L., Milnova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nordb,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,M., Roy,A., Santos,R., Schauer,S., Schnpback,R.,
Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S., Schnpback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thmann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,D., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

TITLE
JOURNAL
COMMENT

Center: Whitehead Institute/MIT Center for Genome Research
 Center code: MIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center Project name: L17240
 Center Clone name: 240_K-1
 ----- Summary Statistics -----
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Assembly quality: 220831 bases at least Q40
 Consensus quality: 221147 bases at least Q30
 Consensus quality: 221206 bases at least Q20
 Insert size: 230000; agarose-1p
 Insert size: 221266; sum-of-ctrls

193021 TTACATGCTTCCTTGGGGCTGAGATGCTGCTCTTACTACATGGCCTA 192972

Rockville, MD 20850, USA
location/qualifiers

Location/Qualifiers

source 1. .926
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="7"

gene
/gene="olfactory receptor MOR251-4P"
/note="potential coding region disrupted by one or more internal stop codons"
/pseudo

BASE COUNT 227 a 209 c 181 g 309 t
ORIGIN

alignment_scores:
Quality: 1354.00 Length: 308
Ratio: 4.544 Gaps: 2
Percent Similarity: 96.753 Percent Identity: 84.091

Alignment_block:
us-09-975-308-9 x AY073900 ..

Align seg 1/1 to: AY073900 from: 1 to: 926

```

1 MetasnHisSerValAlaThrGluPheIleIleuGlyLeuThrLys 17
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4 ATGAACCTTCACGATGTTAGTAGATTATGATGCTGGACTTACTCAAA 53
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 sProGluLeuGlnGlyIleIlePheLeuPheLeuIleValTyrLeu 34
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
54 GTGCAACTTCAGGGAATACATCTTTCATGTTCTTTTATCTACCTTG 103
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 AlaIaPheLeuGlyAsnMetLeuIleIleIleAlaIleTyrSerAsn 50
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
104 TGGCTCTACTGTTGTAATGCTAAATGTTGTCATATCTAATACACC 153
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 ThrIleuHisThrProMetTyrValPheLeuLeuThrIleuAlaVal 67
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154 ACCTTGGCACACCCATGATATCTCTCTGGCCTTGGCTGTGCTGA 203
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 PileIleCysThrThrSerIleIleProLysMetLeuGlyThrMetLeu 84
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
204 CATATCTGCATCAAGCATCATACCCAAATGTTGGAACTAGTAA 253
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 hSerGluAsnThrIleSerTyrAlaGlyCysMetSerGlnLeuPheLeu 100
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
254 CATCAAAAAATTCATTTATATGAGGGGTTGCATGCTCCAGCTCTTT 303
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 PheThrTyrSerLeuGlyAlaGluMetValLeuPheThrThrMetAla 117
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
304 TTCACATGGTCTCTGGGGCTGAGATGGTCTTACTACAAATGGCTTA 353
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 rAspArgTyrValAlaIleCysPheProLeuHisTyrSerThrIleMet 134
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
354 TGACCCGCTATGTGGCCATTTGCTCCACTGCTATAGTACTATTATA 403
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 snHisHisMetCysValAlaLeuLeuSerMetValMetAlaIleAla 150
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
404 ACCACTATACATGTGATGGCTTGTAGCATGTGATGGCTATGCTGTA 453
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 ThrAsnSerTyrValHisThrAlaLeuIleMetArgLeuThrPheCys 167
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454 ACCAATTCCTGGGTGCACACTGCTCATCTGAGGCTGACTTCTGTGG 503
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 yProAsnThrIleAspHisPhePheCysGluIleProProLeuLeu 183
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
504 GCGGAATATATATGACCACTTCTTCTGTAATATACCCCGCTGCTGG 553
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
183 aLeuSerCysSerProValArgIleAsnGluValMetValTyrValAla 200
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
554 TTGTCTCTGATAGTCCCTAGAGATCATGAGTATGATGCTATGCTGG 603
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
200 splIleThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyr 216
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
604 ATATACCCCTGGCTGTGGAGACTTACTACCTGATTTCCCTATGA 653
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
217 PheIleIleValAlaIleLeuArgIleArgThrValGluGlyLysArg 233
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
654 TTTATTATGTCGCTATCTCCGATCCGACACAGAGCAAGCAAGAA 703
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
233 sAlaPheSerThrCysSerSerHisLeuThrValAlaThrLeuTyr 250
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
704 GGCCTTCTTACGTCTCATGACACCTCATGAGTGTGCTTCTACTAT 753
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
250 eProValIleTyrThrTyrIleArgProAlaSerSerTyrThrPheG 266
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
754 CTCCTGTATCTACACCTATATCCAGCTGACATCCAGTATACCTTGA 803
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
267 ArgAspLysValValAlaAlaLeuTyrThrLeuValThrProThr 283
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
804 AAGACAAAGGTGTGCTGCTGCTCATATATCTAGTATACCTTGA 853
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
283 nProMetValTyrSerPheGlnAsnArgIleMetGlnAlaGlyIle 300
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
854 CCCAATAGTATATAGCTTCAGGAACAAAGATGAGTCAAGAAATTA 903
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
300 yValaPheAlaPheLeuLys 306
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
904 AAGTATTTGCATTTTAAAA 923
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
seq_name: gb_pat:AX350633
seq_documentation_block:
LOCUS AX350633 798 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 11 from Patent WO01771177.
ACCESSION AX350633
VERSION AX350633.1 GI:18616201
KEYWORDS
SOURCE
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (sites) Gangolli, E.A., Rastelli, L., Smithson, G., Padigar, M.,
Vernet, C.A., Wolenc, A.R., Casman, S.J., Tcheney, V.T.,
Szekeres, E.S., Gorisse, W., Alsobrook, D.P. and Burgess, C.E.
Novel gpcr-proteins and nucleic acids encoding same
JOURNAL Patent: WO 0177117-A 11 18-OCT-2001;
Curagen Corporation (US)
FEATURES
Location/Qualifiers
source 1..798
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 191 a 202 c 164 g 241 t
ORIGIN

```

alignment_scores:
Quality: 1339.00 Length: 259
Ratio: 5.170 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.228

Alignment_block:
us-09-975-308-9 x AX350633 ..

Align seg 1/1 to: AX350633 from: 1 to: 798

```

49 SerAsnThrLeuHisThrProMetTyrValPheLeuLeuThrIleuAla 65
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3 AACCAACCTTCATACGCGCATGATGTTTCTTCCCTGACAGCTGGCTGT 52
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
65 lValAspIleIleCysThrThrSerIleIleProLysMetLeuGlyThr 82
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
53 TGTGGACATCATCTGCACAAAGCATCATACCGAAGATGCTGGGAGACA 102
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
82 eLeuThrSerGluAsnThrIleSerTyrAlaGlyCysMetSerGlnLeu 98
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

103 TGCATACATAGAAATACATTTCATATGCAGCTGCATGCCACTC 152
99 PheLeuPheThrTrpSerLeuGlyAlaGluMetValLeuPheThrMet 115
153 TTCCTTGTTCACATGTCCTGGAGCGTGAAGATGTTCTCTTCACACCAT 202
115 TAlaTyRAspArgTrpValAlaIleCysPheProLeuHisTySerThr 132
203 GGCTATGACCGCTATGTGGCATTGTTCCTTCCTTCATACAGTACTG 252
132 LeuMetAsnHisMetCysValAlaLeuLeuSerMetValMetAlaIle 148
253 TTTATGACACCATATGTGTGTGCTGCTGCATGAGGTCATGGCATTT 302
149 AlaValThrAsnSerTrpValHisThrAlaLeuIleMetArgLeuThr 165
303 GCAGTCACCAATTCCTGGGTGCACACAGCTTATCATGAGTTGACTTT 352
165 eCysGlyProAsnThrIleAspHisPheCysGluIleProLeu 182
353 CTGTGGCCAAACACCATTTGACCATCTTCTGTGAGATACCCCATTCG 402
182 euAlaLeuSerCysSerProValArgIleAsnGluValMetValTyRVal 198
403 TGGCTTGTCTCTGACCCCTGTAAAGATCATAGGATGATGTTAGTT 452
199 AlaAspIleThrLeuAlaIleGlyAspPheIleLeuThrCysIleSer 215
453 GCTGATATATACCTCGCCATAGGGGACTTATCTTACCTGCATCTCT 502
215 rGlyPheIleIleValAlaIleLeuArgIleArgThrValGluGlyAs 232
503 TGGTTTATCATTTGTCCTATCTCCGATCCGACAGATGAAGAGCAGA 552
232 rGlyAlaPheSerThrCysSerSerHisLeuThrValAlaThrLeuTy 248
553 GGAAGGCTCTTCACATGCTCATCTCATCTCAGAGTGAGCCCTTAC 602
249 TyrSerProValIleTyThrTrpIleArgProAlaSerSerTyThr 265
603 TATTCCTCGTATATCTACACCTATATCCGCCCTGCCTCCAGTATCAT 652
265 eGluAlaGlyAspIleValAlaIleLeuTyThrLeuValThrProThr 282
653 TGAAGACAGACAGGTGTAGCTCATCTTACTCTTGACCTCCACAT 702
282 euAsnProMetValTySerPheGluAsnArgIleMetGluAlaGlyIle 298
703 TAAACCCGATGTGTACAGCTTCAGAAATAGGAGATGCAGCAGGAAT 752
299 ArgIleValPheAlaPheLeuHis 307
753 AGGAAGGTGTTCGATTCTGAAACAC 779

```

seq_name: gb_pat:AX350675

seq_documentation_block:

LOCUS AX350675 663 bp DNA linear PAT 06-FEB-2002
 DEFINITION Sequence 53 from Patent WO0171177.
 ACCESSION AX350675
 VERSION AX350675.1 GI:18616221
 KEYWORDS

SOURCE

ORGANISM

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

Vernet,S., Gangolli,E.A., Rastelli,L., Smithson,G., Padigaru,M.,
 Szekeres,E.S., Gorse,W., Alsobrook,J.P., and Burgess,C.E.

TITLE Novel gpcr-proteins and nucleic acids encoding same
 JOURNAL Patent: WO 0171177-A 53 18-OCT-2001;

Curagen Corporation (US)
 FEATURES Location/Qualifiers
 source 1..663
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 151 a 174 c 134 g 204 t
 ORIGIN

alignment_scores:
 Quality: 1143.00 Length: 221
 Ratio: 5.172 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.548

alignment_block:
 US-09-975-308-9 x AX350675 ..

Align seg 1/1 to: AX350675 from: 1 to: 663

```

63 LeuAlaValAlaAspIleIleCysThrThrSerIleLeuProIleMetLe 79
1 CTGGCTGTGTGGACATCATCTGCACAAAGACATCATACCGAAGATGCT 50
79 uGlyThrMetLeuThrSerGluAsnThrIleSerTyRAlaGlyCysMet 96
51 GGGGACCATCTCAACATCAGAAATACATTTCATATGACGGCTCATGT 100
96 eGluLeuPheLeuPheThrTrpSerLeuGlyAlaGluMetValLeuPhe 112
101 CCAGGCTCTCTGTTCACATGCTCTGTGGAGCTGAGATGTTCTCTTC 150
113 ThrThrMetValTyRAspArgTrpValAlaIleCysPheProLeuHis 129
151 ACCACCATGCGCTATGACCGCTATGTGGCATTGTTCCTCTTCATTA 200
129 rSerThrIleMetAsnHisMetCysValAlaLeuLeuSerMetValM 146
201 CACTAGCTATATGACACCATATGTGTGAGCTTGCTGATGATGATGTA 250
146 eAlaIleAlaValAlaThrAsnSerTrpValHisThrAlaLeuIleMet 162
251 TGGCTATGTGACACCAATTCCTGGGTGCACACACTTATCATCAGAGG 300
163 LeuThrPheCysGlyProAsnThrIleAspHisPhePheCysGluIleP 179
301 TTGACTTTCTGTGGCCAAACACCATTTGACCACTTCTTCTGTAGATAC 350
179 oProLeuLeuAlaLeuSerCysSerProValArgIleAsnGluValMet 196
351 CCCATTGCTGGCTTGTGCTGTAGCCCTGTAAAGATCAATGAGGTGATG 400
196 aTyRValAlaAspIleThrLeuAlaIleGlyAspPheIleLeuThrCys 212
401 TGTATGTGTGATATATACCTGCGCATAGGGACCTTATTTACTCTGC 450
213 IleSerTyGlyPheIleIleValAlaIleLeuArgIleArgThrValG 229
451 ATTCTCTATGTTTATCATTTGCTATTTCTCGATATCCGACAGTAGA 500
229 uGlyTyRArgIleAlaPheSerThrCysSerSerHisLeuThrValAla 246
501 AGGCAAGAGAGAGCGCTTGCACATGCTCATCTCATCTCAGAGTGTGA 550
246 hLeuTyRTrpSerProValIleTyThrTrpIleArgProAlaSerSer 262
551 CCTTTACTATTCTCCTGTAAATCTACACCTATATCCGCCCTGCTTCAC 600
263 TyRThrPheGluArgAspIleValAlaIleLeuTyThrLeuValTh 279
601 TATACATTTGAAAGACAAAGGTGTAGCTGCATCTTATCTTGTGAC 650
279 rProThrLeuAsn 283
|||||

```

651 TCCACATTAAC 663

seq_name: gb_htg:AC101272

seq_documentation_block:

AC101272 59688 bp DNA linear HTG 23-NOV-2001

DEFINITION Mus musculus clone RP23-10J20, LOW-PASS SEQUENCE SAMPLING.

AC101272

AC101272.1 GI:17060047

HTG: HTGS_PHASE0.

SOURCE

house mouse

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 59688)

Mus musculus, clone RP23-10J20

Unpublished

2 (bases 1 to 59688)

Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,

Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B.,

Choquel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,

Cooke, P., DeArrelano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,

Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K.,

Lamaze, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,

Maclean, C., MacDonald, P., Major, J., Margulis, N., Matthews, C.,

McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,

Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,

Norman, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,

Oliver, J., Peterson, K., Phukhang, P., Pierre, N., Pollara, V.,

Raymond, C., Retta, R., Ribick, M., Riley, R., Rise, C., Rogov, P.,

Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,

Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,

Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,

Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L16346

Center clone name: 101_J_20

* NOTE: This record contains 76 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely for

* arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will

* be preserved.

* 1 679: contig of 679 bp in length

* 680 779: gap of 100 bp

* 780 1489: contig of 710 bp in length

* 1490 1589: gap of 100 bp

* 1590 2263: contig of 674 bp in length

* 2264 2363: gap of 100 bp

2364 3054: contig of 691 bp in length
3055 3154: gap of 100 bp
3155 3839: contig of 685 bp in length
3840 3939: gap of 100 bp
3940 4630: contig of 691 bp in length
4631 4730: gap of 100 bp
4731 5401: contig of 671 bp in length
5402 5501: gap of 100 bp
5502 6164: contig of 663 bp in length
6165 6264: gap of 100 bp
6265 6949: contig of 665 bp in length
6950 7049: gap of 100 bp
7050 7728: contig of 679 bp in length
7729 7828: gap of 100 bp
7829 8516: contig of 688 bp in length
8517 8616: gap of 100 bp
8617 9325: contig of 709 bp in length
9326 9425: gap of 100 bp
9426 10127: contig of 702 bp in length
10128 10227: gap of 100 bp
10228 10924: contig of 697 bp in length
10925 11024: gap of 100 bp
11025 11712: contig of 688 bp in length
11713 11812: gap of 100 bp
11813 12481: contig of 669 bp in length
12482 12581: gap of 100 bp
12582 13254: contig of 673 bp in length
13255 13354: gap of 100 bp
13355 14033: contig of 679 bp in length
14034 14133: gap of 100 bp
14134 14816: contig of 683 bp in length
14817 14916: gap of 100 bp
14917 15597: contig of 681 bp in length
15598 15697: gap of 100 bp
15698 16388: contig of 691 bp in length
16389 16488: gap of 100 bp
16489 17234: contig of 746 bp in length
17235 17334: gap of 100 bp
17335 18009: contig of 675 bp in length
18010 18109: gap of 100 bp
18110 18802: contig of 693 bp in length
18803 18902: gap of 100 bp
18903 19608: contig of 706 bp in length
19609 19708: gap of 100 bp
19709 20385: contig of 677 bp in length
20386 20485: gap of 100 bp
20486 21178: contig of 693 bp in length
21179 21278: gap of 100 bp
21279 21942: contig of 664 bp in length
21943 22042: gap of 100 bp
22043 22714: contig of 672 bp in length
22715 22814: gap of 100 bp
22815 23486: contig of 672 bp in length
23487 23586: gap of 100 bp
23587 24261: contig of 675 bp in length
24262 24361: gap of 100 bp
24362 25046: contig of 685 bp in length
25047 25146: gap of 100 bp
25147 25807: contig of 661 bp in length
25808 25907: gap of 100 bp
25908 26596: contig of 689 bp in length
26597 26696: gap of 100 bp
26697 27406: contig of 710 bp in length
27407 27506: gap of 100 bp
27507 28223: contig of 717 bp in length
28224 28323: gap of 100 bp
28324 29026: contig of 703 bp in length
29027 29126: gap of 100 bp
29127 29814: contig of 688 bp in length
29815 29914: gap of 100 bp
29915 30588: contig of 674 bp in length
30589 30688: gap of 100 bp
30689 31361: contig of 673 bp in length

[illegible]

Quality:	964.00	Length:	213
Ratio:	4.657	Gaps:	0
Percent Similarity:	97.183	Percent Identity:	84.977

```
alignment_block:
US-09-975-308-9 x AC101272 ..
```

Align seg 1/1 to: AC101272 from: 1 to: 596888

14 LeuThrLysLysProGluLeuGlnGlyIleIlePheLeuPheLeuI 30
|||||::||| |||||||||:::||||: |||||:
25170 CTTACTCAAAAGCTGAACTTCAGGSAATACTTTTCATGTTTTCTTTT 25219

```

30  eValIytleValaIaPheLengIySmetLeuIleIleIaIaIygr  47
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
25220  TACTATCCTTGtGGCTCACTGtGTAATAGCTAAATGtGTGGCCATGA  25269
   47  IeYtSerAsnThrLeuHisIsthPrMetIyValPheLeuLeuThrLeu  63
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
25270  TGTATTAACACACCTTGCACACACCACCATGATATCTCCCTTGCGGCTTG  25319
   64  AlAValAlAspIleIleCysThrThrSerIleIleProIySmetLeuGl  80
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
25320  GCGTGTGGACATAACTGACACTACACACATCATGCCAAAATGTTGGG  25369
   80  yThrMetLeuThrSerGIuAsnThrIleSerYtAlaGIyCysMetSerG  97
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
25370  AACATGTTAAACATCAAAAATTCATTTTCATATGGGGGTTGCATGTCCC  25419
   97  ILeuPheLeuPheThrTrpSerLeuGIyAlaGIuMetValLeuPheThr  113
   AGCTCTCTTTTTCACATGtGTCCTTGGGGGCTGAGATGGTCTTTACT  25469
   114  ThrMetAlaYtAspArgTyrValAlaIleCysPheProLeuHisTyrSe  130
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
25470  ACATAGCCCTATGACCGCTATGtGGCCATTTGCTTCCACTTGCTATAG  25519
   130  rThrIleMetAsnHisHisMetCysValAlaLeuLeuSerMetValMetA  147
   TACTATTTATGACACCATATACATGtGTAGGCTTGCTTACCATTTGTATG  25569
   25520  TACTATTTATGACACCATATACATGtGTAGGCTTGCTTACCATTTGTATG  25569
   147  IAlaIeAlaValThrAsnSerTrpValHisThrAlaLeuIleMetArgLeu  163
   25570  CTAATCTGCTAAACCAATTCCTGGGTCACACACTGTCTCATTTCTGACGG  25619
   164  ThrPheCysGIyProAsnThrIleAspHisPhePheCysGIuIleProPr  180
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
25620  ACTTGTGtGGGCGGAAATTAATGACCACTTTTGTGAATTAACCCCC  25669
   180  oLeuLeuAlaLeuSerCysSerProValAlArgIleAsnGIuValMetValTr  197
   25670  GCGTCTGCTTGTGCTGCTAGTCTGTAAGACATGAGGTAGTGTGCTG  25719
   197  yValAlaAspIleThrLeuAlaIleGIyAspPheIleLeuThrCysIle  213
   25720  ATGTTCTGTAATACACCTGCGTGTGGAGACTTACTCTAACCTGCATTT  25769
   214  SerTyrGIyPheIleIleValAlaIleLeuArgIleArg  226
   25770  TCCTATGATTTTATTATTCGTGCTATTCTCCGCACTCCG  25808
   seq_name: gb_to:AY073135

seq_documentation_block:
LOCUS      AY073135                915 bp    DNA       linear   ROD 04-FEB-2002
DEFINITION Mus musculus olfactory receptor MOR251-3 gene, complete cds.
ACCESSION  AY073135
VERSION    AY073135.1  GI:18479567
KEYWORDS
SOURCE
ORGANISM   house mouse.
            Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murineae; Mus.
REFERENCE  1 (bases 1 to 915)
            Zhang,X. and Firestein,S.
            The olfactory receptor gene superfamily of the mouse
            Nat. Neurosci. 5 (2), 124-133 (2002)
PUBMED     11802173
REFERENCE  2 (bases 1 to 915)
            Adams,M.
            Direct Submission
            Submitted (11-JAN-2002) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
FEATURES
            Location/Qualifiers
            1..915

```

```

mRNA
CDS
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /chromosome="7"
    /product="olfactory receptor MOR251-3"
    1. 915
        /codon_start=1
        /evidence=not experimental
        /product="olfactory receptor MOR251-3"
        /protein_id="AI60798.1"
        /db_xref="GI:18479568"
        /translation="MNGTLVTEFLILFSDFMRHRIILFSLFQIYVAASGNLLIV
        TISASPLHPRMYFEVLVAIVBDILSTLPLKLDVWGRTISYGCMAQLEFFW
        SLAEVLEFSAAMDREVAACPLHYSITWMPRCAPLAGIWAISLNTINSGLV
        RLPFCSNVAWEHFCEIPLRLKLSCAPTOENAAAFADVLAVGNFVLIISGLV
        ASILKISAEKQRAFSCTGSAHLIVMYSTVIYTIIRSSYSLSNKKRVSLIYS
        VAPLNLILYLRKKDVKVALRRLFSC"

BASE COUNT      185 a      279 c      198 g      253 t
ORIGIN

alignment_scores:
    Quality: 955.50      Length: 303
    Ratio: 3.552      Gaps: 1
Percent Similarity: 88.779      Percent Identity: 55.776

Alignment_block:
US-09-975-308-9 x AY073135 ..

Align seg 1/1 to: AY073135 from: 1 to: 915

1 MetAsnHisSerValValThrGluPheIleIleLeuGlyLeuThrIlyslu 17
||||| .....|
1 ATGAAATGGGACACTGGTCACTGAGTTCCTCATCTGGAGATTCCAGATAT 50
17 sproGluLeuGlnGlyIleIlePheLeuPhePheLeuIleValIlyrLeu 34
|||||.....|
51 GCCTCACTTCGGATCACTACTTTCTCCACACTTCCTATAGCCTATACATGG 100
34 aAlaPheLeuGlyAsnMetLeuIleIleIleAlaIlySerAsn 50
|||||.....|
101 CTCACATCTGCAGAAACCTGCTATTATGGTGACAAATCACTGCAGGCCA 150
51 ThrLeuHisThrProMetIlyValPheLeuLeuThrIleuAlaValas 67
|||||.....|
151 ACCCTGCATACCCCTATGACTCTTCTCTGATCACTGAGCTGTGGTGA 200
67 PileIleCysThrThrSerIleIleProIlyMetLeuGlyThrMetLeu 84
|||||.....|
201 CACTCTTGGACCTCCACACATCCATACCAAGCTCATGGACATGGTGG. 249
84 hIserGluAsnThrIleSerThrIalGlyCysMetSerGlnLeuPheLeu 100
|||||.....|
250 ..GGGGGAAGACCATCTCTACGGGGGGCTGCATGGCCAGCTTCTTC 297
101 PheThrTrpSerLeuGlyAlaGluMetValLeuPheThrThrMetAlaIy 117
298 TTTCATATGTCACACTGGGGGTGAGACTTCTACTCTTCTCAGCTATGGCCTA 347
117 rAsrArgTyrValAlaIleCysPheProLeuHisTyrSerThrIleMetA 134
|||||.....|
348 TGACCCGCTTGTGGCATCTGCTGCCCTCTCATATTATAGTACCTGAGTGG 397
134 snHisHisMetCysValAlaLeuLeuSerMetValMetAlaIleAlaVal 150
|||||.....|
398 GCCCAGAGGTGTGCATTTCTGGCTGGCATTTGCTGGGCTATACAGCCTG 447
151 ThrAsnSerTrpValHisThrAlaLeuIleMetArgLeuThrPheCysGl 167
|||||.....|
448 ACTAACACACACATTAACCTCAAGCCTGGTGGCTCATCAACATTTGGAG 497
167 yProAsnThrIleAsnHisPheCysGluIleProProLeuLeuAlaVal 184
|||||.....|

```

```

498 CTCCAAATGCGTTGAGCAGCTTCTCTGTGGANATTCGCCCACTGTTGAAGC 547
184 euserCysSerProValArgIleasnIuValMetValTyrValAlaasp 200
548 TCTCTTGTCGTCACAAACAATTGAACAGGCGCATGGCTTTCGTCGACAGT 597
201 IlethreunAlaIleGlyAspPheIleuThrcysIleSerTyrGlyPh 217
598 GTGTTCTCGGCTGAGGGAAGCTTCTCTGTGATCATCTCTCTGATGGCTT 647
217 eileIleValAlaIleuArgIleArgTyrValIleuGlyAsrGlysa 234
648 TATGTGGCCAGCACTCGAAGATCCGCTCAGCTGAGGGAAGCAACGAG 697
234 IapheSerThrcysSerSerThsIeuThrValIleuThleuTyrTyrSer 250
698 CCTTCTCCACCTGCTCTGCACACCTCAGTCGIGTGATCACCATGACTACTCC 747
251 ProValIleTyrThrTyrIleArgProAlaSerSerTyrThrPheGluAr 267
748 ACTGCATCTACACCTACATTCGCCCTTCAATCCACTACCTCAACAA 797
267 GAspIysValValAlaIleuTyrThleuValhrProThrIeuAsp 284
798 GGACAGAGTGCTGCATCATCTACACCTCAGTGGCAGCCACCTTGAGC 847
284 IometValTyrSerPheGlnAsnArgIuMetGlnAlaGlyIleArgIys 300
848 CTCCTATCTACACTCGAAGATTAAGATGTCAAGTTGACATCCGGAGA 897
301 ValPheAla 303
898 CTTTCTCTCC 906
seq_name: gb AY074180
seq_documentation_block:
LOCUS AY074180 916 bp DNA linear ROD 04-FEB-2002
DEFINITION Mus musculus olfactory receptor MOR251-5 pseudogene, partial
sequence.
ACCESSION AY074180
VERSION AY074180.1 GI:18481342
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 916)
Zhang, X. and Firestein, S.
The olfactory receptor gene superfamily of the mouse
Nat. Neurosci. 5 (2), 124-133 (2002)
JOURNAL PUBMED 11802173
REFERENCE
2 (bases 1 to 916)
Adams, M.
Direct Submission
Submitted (11-JAN-2002) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
FEATURES
source
1..916
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="7"
<1..>916
/genes="Olfactory receptor MOR251-5"
/notes="potential coding region disrupted by one or more
internal stop codons"
/pseudo
/evidence=not_experimental
BASE COUNT 185 a 287 c 200 g 244 t
ORIGIN
alignment_scores:

```

seq_name: gb_htg:AC094870

seq_documentation_block:

LOCUS AC094870 112730 bp DNA linear HTG 20-DEC-2001
DEFINITION Rattus norvegicus clone CH230-5N18, *** SEQUENCING IN PROGRESS ***
52 unordered pieces.
ACCESSION AC094870
VERSION AC094870.2 GI:17941654
KEYWORDS HTGS, HTGS_PHAISEI.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 112730)
Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Blmage,K., Blankenburg,K., Donnan,D., Bouck,J., Bowie,S., Briveva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Butcher,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C., Cleeland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denun,A.L., Ding,Y., Dinh,H.H., Douhaite,K.J., Draper,H., Duncan-Rocha,S., Dublin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.Y., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsí,F., Howard,S., Huber,J., Hulysk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Koryah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J.Z., Lichtarge,O., Lieu,C., Liu,C., Liu,W., Lounsgaed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meddor,M., Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbet,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,S., Newbott,K., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenko,S.S., Ogihara,N., Okunoye,G., Orangunye,N., Oyedero,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojcs,A., Rojbockan,T., Rolfe,M., Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shooshitari,N., Slison,I., Sodergren,E., Sonakke,T., Sparks,A., Stanley,H., Stone,H., Sultan,A., Swatek,A., Tabori,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,Y., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wlezlitz,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D., Weinstein,G. and Gbbbs,R.

TITLE Unpublished
JOURNAL Direct Submission
REFERENCE 2 (bases 1 to 112730)
Worley,K.C.
AUTHORS Direct Submission
TITLE Submitted (15-SEP-2001) Human Genome Sequencing Center, Department

COMMENT

Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15624706.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GBPG

Center clone name: CH230-5N18

----- Summary Statistics

Assembly program: Phrap; version 0.990329First call to

findphraplist

Consensus quality: 96697 bases at least Q40

Consensus quality: 104847 bases at least Q30

Consensus quality: 112341 bases at least Q20

Estimated insert size: 92352; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 1.1x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 52 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

```

1      6916: contig of 6916 bp in length
*      6917      7016: gap of unknown length
*      7017      11089: contig of 4073 bp in length
*      11090      11189: gap of unknown length
*      11190      14317: contig of 3328 bp in length
*      14318      14617: gap of unknown length
*      14618      17992: contig of 3375 bp in length
*      17993      18092: gap of unknown length
*      18093      21658: contig of 3566 bp in length
*      21659      21758: gap of unknown length
*      21759      25307: contig of 3549 bp in length
*      25308      25407: gap of unknown length
*      25408      27818: contig of 2411 bp in length
*      27819      27918: gap of unknown length
*      27919      31526: contig of 3608 bp in length
*      31527      31626: gap of unknown length
*      31627      33967: contig of 2341 bp in length
*      33968      34067: gap of unknown length
*      34068      36642: contig of 2575 bp in length
*      36643      36742: gap of unknown length
*      36743      39205: contig of 2463 bp in length
*      39206      39305: gap of unknown length
*      39306      41281: contig of 1976 bp in length
*      41282      41381: gap of unknown length
*      41382      45065: contig of 3684 bp in length
*      45066      45165: gap of unknown length
*      45166      47143: contig of 1978 bp in length
*      47144      47243: gap of unknown length
*      47244      48499: contig of 1256 bp in length
*      48499      48599: gap of unknown length
*      48599      50055: contig of 1456 bp in length
*      50055      50155: gap of unknown length
*      50156      53530: contig of 3375 bp in length
*      53531      53630: gap of unknown length
*      53631      55445: contig of 1815 bp in length
*      55446      55545: gap of unknown length
*      55546      58015: contig of 2470 bp in length
*      58015      58115: gap of unknown length
*      58116      59612: contig of 1497 bp in length
*      59612      59712: gap of unknown length
*      59713      61702: contig of 1990 bp in length
*      61702      61802: gap of unknown length
*      61802      63157: contig of 1355 bp in length
*      63157

```

```

*      63158      63257: gap of unknown length
*      63258      64623: contig of 1366 bp in length
*      64624      64723: gap of unknown length
*      64724      65766: contig of 1043 bp in length
*      65767      65866: gap of unknown length
*      65867      67869: contig of 2003 bp in length
*      67870      67969: gap of unknown length
*      67970      69560: contig of 1591 bp in length
*      69561      69660: gap of unknown length
*      69661      71980: contig of 2320 bp in length
*      71981      72080: gap of unknown length
*      72081      73657: contig of 1577 bp in length
*      73658      73757: gap of unknown length
*      73758      75695: contig of 1938 bp in length
*      75696      75795: gap of unknown length
*      75796      78395: contig of 2600 bp in length
*      78396      78495: gap of unknown length
*      78496      79707: contig of 1212 bp in length
*      79708      79807: gap of unknown length
*      79808      81374: contig of 1567 bp in length
*      81375      81474: gap of unknown length
*      81475      83163: contig of 1689 bp in length
*      83164      83263: gap of unknown length
*      83264      85474: contig of 2211 bp in length
*      85475      85574: gap of unknown length
*      85575      87264: contig of 1630 bp in length
*      87265      87364: gap of unknown length
*      87365      88370: contig of 1006 bp in length
*      88371      88470: gap of unknown length
*      88471      89719: contig of 1249 bp in length
*      89720      89819: gap of unknown length
*      89820      91531: contig of 1712 bp in length
*      91532      91631: gap of unknown length
*      91632      92853: contig of 1222 bp in length
*      92854      92953: gap of unknown length
*      92954      95061: contig of 2108 bp in length
*      95062      95161: gap of unknown length
*      95162      97151: contig of 1990 bp in length
*      97152      97251: gap of unknown length
*      97252      98827: contig of 1576 bp in length
*      98828      98927: gap of unknown length
*      98929      100300: contig of 1373 bp in length
*      100301      100400: gap of unknown length
*      100401      101467: contig of 1067 bp in length
*      101468      101567: gap of unknown length
*      101568      102772: contig of 1205 bp in length
*      102773      102872: gap of unknown length
*      102873      104128: contig of 1236 bp in length
*      104129      104228: gap of unknown length
*      104229      105514: contig of 1286 bp in length
*      105515      105614: gap of unknown length
*      105615      106797: contig of 1183 bp in length
*      106798      106897: gap of unknown length
*      106898      108212: contig of 1315 bp in length
*      108213      108312: gap of unknown length
*      108313      109627: contig of 1315 bp in length
*      109628      109727: gap of unknown length
*      109728      110921: contig of 1194 bp in length
*      110922      111022: gap of unknown length
*      111022      112730: contig of 1709 bp in length.

```

FEATURES

source

1. 112730
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-5N18"

BASE COUNT 30919 a 22306 c 22742 g 31565 t 5198 others
ORIGIN

alignment_scores: 935.00 Length: 303
Quality: 3.476 Gaps: 2
Ratio: 88.779 Percent Identity: 55.446
Percent Similarity: 88.779

Align seg 1/1 to reverse of: AC094870 from: 1 to: 112730

284 r0meValTyIseRphneGlnAsnAGlMecInlaGlyLeaYlys 300
|||:::|||||:::|||||:::|||||:::|||||:::|||||
95497 CCCTATCTACACACTCTGAGGAACAAGAATGTCAAACTTCACCTCCGAAA 95448

301 ValPheala 303
:::|||||
95447 CTTCTGTCC 95439

seq_name: gb_htg:AC094700

seq_documentation_block:
LOCUS 197438 bp DNA linear HTG-20-DEC-2001
DEFINITION Rattus norvegicus clone CH230-5F7, *** SEQUENCING IN PROGRESS ***,
66 unordered pieces.
ACCESSION AC094700
VERSION AC094700.4 GI:17941479
KEYWORDS HTG; HTGS-PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
1 (Phases 1 to 197438)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunga,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blumge,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowles,E., Brieva-McK., Brown,E., Brown,M., Bryant,N.P., Bunney,C.,
Butch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carrott,T.F.,
Cartier,M., Cavazos,S.R., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dalborne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Dean,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Forster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunnaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Haves,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsl,F., Howard,S., Huber,J., Huliy,S., Hunne,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Matlindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,K.,
Mei,G., Melker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nnokenkwo,S.,
Oguh,M., Okunnu,G., Oragunye,N., Oyedelo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quilis,M., Ren,Y., Rives,M., Rojas,A., Rojudoacan,I., Rolfe,M.,
Ruiz,S., Saverly,G., Scherer,S., Scott,Z., Shen,H., Shooshbari,N.,
Sisson,I., Sodergren,E., Sonstke,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabdor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Tellrod,B., Thoms,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Unpublished
2 (Phases 1 to 197438)
Worley,K.C.
Direct Submission
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17062158.
Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: G8W
Center clone name: CH230-5F7
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329first call to findhaplist
Consensus quality: 163315 bases at least Q40
Consensus quality: 177813 bases at least Q30
Consensus quality: 189359 bases at least Q20
Estimated insert size: 163704; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 1.8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 96 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
6380: contig of 6380 bp in length
6381 6480: gap of unknown length
6481 11235: contig of 4755 bp in length
11335 11335: gap of unknown length
11336 14082: contig of 2747 bp in length
14083 14182: gap of unknown length
14183 18911: contig of 4729 bp in length
18911 22084: contig of 3073 bp in length
22084 22184: gap of unknown length
22184 26199: contig of 4015 bp in length
26199 26200: gap of unknown length
26200 28568: contig of 2269 bp in length
28568 28569: gap of unknown length
28569 31639: contig of 2971 bp in length
31639 31739: gap of unknown length
31739 33776: contig of 2037 bp in length
33776 33876: gap of unknown length
33876 37642: contig of 3766 bp in length
37642 37742: gap of unknown length
37742 39937: contig of 2195 bp in length
39937 40037: gap of unknown length
40037 43330: contig of 3293 bp in length
43330 43430: gap of unknown length
43430 46413: contig of 2983 bp in length
46413 46513: gap of unknown length
46513 49924: contig of 3411 bp in length
49924 50024: gap of unknown length
50024 52777: contig of 2753 bp in length
52777 52877: gap of unknown length
52877 55322: contig of 2445 bp in length
55322 55422: gap of unknown length
55422 57118: contig of 1696 bp in length
57118 57219: gap of unknown length
57219 59661: contig of 2443 bp in length
59661 59761: gap of unknown length
59761 62279: contig of 2518 bp in length
62279 62379: gap of unknown length
62379 65763: contig of 3384 bp in length
65763 65863: gap of unknown length
65863 68840: contig of 2977 bp in length
68840 68940: gap of unknown length
68940 70763: contig of 1823 bp in length
70763 70863: gap of unknown length
70863 73785: contig of 2922 bp in length
73785 73885: gap of unknown length

73886 76368: contig of 2483 bp in length
76368 76468: gap of unknown length
76468 79049: contig of 2581 bp in length
79049 79149: gap of unknown length
79149 82088: contig of 2939 bp in length
82088 82188: gap of unknown length
82188 84182: contig of 1994 bp in length
84182 84283: gap of unknown length
84283 87126: contig of 2843 bp in length
87126 87225: gap of unknown length
87225 88998: contig of 1773 bp in length
88998 89098: gap of unknown length
89098 92016: contig of 2917 bp in length
92016 92115: gap of unknown length
92115 94512: contig of 2397 bp in length
94512 94613: gap of unknown length
94613 96876: contig of 2264 bp in length
96876 96976: gap of unknown length
96976 98801: contig of 1825 bp in length
98801 98901: gap of unknown length
98901 100728: contig of 1827 bp in length
100728 100828: gap of unknown length
100828 102598: contig of 1770 bp in length
102598 102698: gap of unknown length
102698 104252: contig of 1554 bp in length
104252 104352: gap of unknown length
104352 105818: contig of 1466 bp in length
105818 105918: gap of unknown length
105918 107016: contig of 1098 bp in length
107016 107116: gap of unknown length
107116 108700: contig of 1584 bp in length
108700 108800: gap of unknown length
108800 110130: contig of 1330 bp in length
110130 110230: gap of unknown length
110230 111400: contig of 1170 bp in length
111400 111500: gap of unknown length
111500 113819: contig of 2319 bp in length
113819 113919: gap of unknown length
113919 115967: contig of 2048 bp in length
115967 116067: gap of unknown length
116067 117164: contig of 1097 bp in length
117164 117264: gap of unknown length
117264 118813: contig of 1549 bp in length
118813 120583: gap of unknown length
120583 120683: contig of 1670 bp in length
120683 122741: contig of 2058 bp in length
122741 122841: gap of unknown length
122841 125104: contig of 2263 bp in length
125104 125204: gap of unknown length
125204 127018: contig of 1814 bp in length
127018 127118: gap of unknown length
127118 128590: contig of 1472 bp in length
128590 128690: gap of unknown length
128690 129819: contig of 1129 bp in length
129819 129919: gap of unknown length
129919 131056: contig of 1137 bp in length
131056 131156: gap of unknown length
131156 133124: contig of 1968 bp in length
133124 133224: gap of unknown length
133224 135098: contig of 1874 bp in length
135098 135198: gap of unknown length
135198 136613: contig of 1415 bp in length
136613 136713: gap of unknown length
136713 138549: contig of 1836 bp in length
138549 138649: gap of unknown length
138649 138550

alignment_scores: quality: 935.00 length: 303
 ratio: 3.476 gaps: 2
Percent Similarity: 86.779 Percent Identity: 55.446
alignment_block:

US-09-975-308-9 x AC094700/rev ..

Align seg 1/1 to reverse of: AC094700 from: 1 to: 197438

```

1 MetasnhiservalValThrgluPhlelleleuGlyLeuThrylsly 17
||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96743 ATGATGGGACACTGGTCACTGATGCTCATCTCGGATTCAGAAAT 96694

17 sprogileuGlylleleuPheleuPheleuIleValtyrleuV 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96693 GCCTACCTTCGG...GTACCACTTCACCTTCCTTCCTATACATGG 96647

34 alAlapheluGlyasmetleuIlelleleuIleValtyrsetAsn 50
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96646 CTGCATTCAGAACTGCTCATTTAGTGAGTCACTGCCAGCCCA 96597

51 ThleuHIsThrPrometyrValPheleuLeuThreValValas 67
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96596 GCCTGCATACCCCTATGTACTTCCTGCTGCACTTACCTGCTGGA 96547

67 pIlelleCysThrThrsertlelleleuPolysmetleuGlyThmetleut 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96546 CATCCCTTCGACCTCCACCACTCCACCACTCTGCATCATGTGTA 96498

84 hrsergluAsnThrIleSertyrAlaGlyCysmetsergluPheleu 100
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96497 .GGGGAGAGACCATCTCTTATGCTGCTGCATGGCCCGCTTCTTC 96450

101 PheThrTrpSerleuGlyAlaIleuMetValleuPheThrThrMetAlaTy 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96449 TTCACATGTCATGGGGCAGAGCTTCCTCTCTCACACTATGGCTTA 96400

117 rAsparGlyValAlaIleCysPheProleuHIsTySerThrIleMetAla 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96399 TGACCTTCCTGGCAATGCTGCTGCCCTGACATAGTACCTGATGG 96350

134 snIshIsMetCysValAlaIleuSerMetValMetAlaIleAlaVal 150
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96349 GCCCCAGGGTGTGCTCTGCTGGCTGCTCTGCTGCTCATCAGCTG 96300

151 ThrAsnSerTrpValHIsThrAlaIleuIleMetArgLeuThrPheCysgl 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96299 ACTAAGACCACTGTGAACCAACCTGTGCTCATCTACCATTCGACG 96250

167 yProAsnThrIleAspHIsPhePheCysgluIleProProleuLeuAla 184
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96249 CTCCATGTGATAGACACTCTCTGTGTGATGCCCACTGTAAAGC 96200

184 euSerCysSerProValArgIleAsnGluValMetValTyValAlaAsp 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96199 TCTCTGTGTCCACACCAATGACAGAGTTATGGCTTGTGTCAGAT 96150

201 IleThrLeuAlaIleGlyAspPheIleuThrCysIleSertyrGlyPh 217
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96149 GTCTTCTGCTAGGAACTCTCTGTGATCATCTCTCTCATGTGCTT 96100

217 eIleIleValAlaIleleuArgIleArgThrValGluGlyAsnGlysa 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96099 TATGTGACAGATATCTGAGATAGCTCACTAGGAGGCAACCAAGAG 96050

234 laPheSerThrCysSerSerHIsLeuThrValAlaThrLeuTyTySer 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96049 CCTTCTACCTGTCTGTGCACTCTTGTGTGTCACCATGTACTACTCC 96000

251 ProValIleTyThrTyrlleArgProAlaSerSertyrThrPheGluAr 267
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95999 ACTATATCTACACTACATTCGCTGATCCAGCTACATCATTAACAA 95950

267 gAspLysValAlaIleleuTyThrLeuValIleProThrLeuAsn 284
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95949 GGCAGAGGTGGTGCATCTACACTGCGTGGCAACCCACCTTAACC 95900

284 rMetValTySerPheGlnAsnArgIleuMetGlnAlaGlyIleArglys 300

```

95899 CCCATCTACACTCTGAGCAACAGATGTCAAGTTCACCTCGGAAA 95850

301 ValPheAla 303

95849 CTCTGTCTCC 95841

seg_name: gb_hhg:AC096461

seg_documentation_block:

LOCUS AC096461

DEFINITION Rattus norvegicus clone CH230-58J17, *** SEQUENCING IN PROGRESS

ACCESSION AC096461.2 GI:17955456

VERSION HTG: HTGS_PHASE1.

KEYWORDS Norway rat.

SOURCE Rattus norvegicus

ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 204143)

AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., All-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barabara,J.,

Benton,J., Blamag,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,

Burch,P., Burke,C., Burrell,K.L., Byrd,N.C., Catron,T.F.,

Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn,A.L., Ding,Y., Dinh,H.H., Donthwaite,K.J., Draper,H.,

Dugan-Rocha,S., Durbin,K.J., Barnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Escoto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,

Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,

Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,

Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,

Hollins,B., Homs,F., Howard,S., Huber,J., Hulky,S., Hume,J.,

Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,

Joudah,S., Karlsson,E., Kelly,S., Khan,D., King,L., Korvah,J.,

Kovar,C., Kratovic,D., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,

Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,

Louiseged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Lune,R.,

Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,

Martinez,E., Massey,E., Mayhew,E., McLeod,M.P., Meador,M.,

Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,

Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,

Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenwo,S.,

Ogulu,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,

Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,

Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubenok,I., Rolfe,M.,

Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostlati,N.,

Slisom,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,

Stone,H., Sutton,A., Syatke,A., Taber,P., Tameris,A., Tameris,K.,

Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,

Thomas,S., Uemari,K., Vasquez,L., Vera,Y., Villalon,D., Vinson,R.,

Wall,R., Wang,S., Ward-Moore,S., Warren,K., Washington,C.,

Watlington,S., Williams,G., Williamson,A., Wlezyg,R., Woodan,S.,

Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,

Weinstock,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 204143)

Worley,K.C.

Direct Submission

Submitted (17-SEP-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced gi:15628089.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

COMMENT

JOURNAL

AUTHORS

REFERENCE

TITLE

Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GFSC
Center clone name: CH230-58J17
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329First call to
findphrapblast
Consensus quality: 156802 bases at least Q40
Consensus quality: 170292 bases at least Q20
Consensus quality: 181161 bases at least Q20
Estimated insert size: 177160; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 62 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1	9619:	contig of 9619 bp in length
9620	9719:	gap of unknown length
9720	18930:	contig of 9211 bp in length
18931	19030:	gap of unknown length
19031	27646:	contig of 8616 bp in length
27647	27746:	gap of unknown length
27747	33370:	contig of 5624 bp in length
33371	33470:	gap of unknown length
33471	41116:	contig of 7646 bp in length
41117	41216:	gap of unknown length
41217	48276:	contig of 7060 bp in length
48277	48376:	gap of unknown length
48377	54436:	contig of 6060 bp in length
54437	54536:	gap of unknown length
54537	61966:	contig of 7430 bp in length
61967	62066:	gap of unknown length
62067	66405:	contig of 4333 bp in length
66406	66505:	gap of unknown length
66506	73151:	contig of 6646 bp in length
73152	73251:	gap of unknown length
73252	79129:	contig of 5878 bp in length
79130	79229:	gap of unknown length
79230	85275:	contig of 6046 bp in length
85276	85375:	gap of unknown length
85376	90469:	contig of 5094 bp in length
90470	90569:	gap of unknown length
90570	94546:	contig of 3977 bp in length
94547	94646:	gap of unknown length
94647	99437:	contig of 4791 bp in length
99438	99537:	gap of unknown length
99538	102811:	contig of 3274 bp in length
102812	102911:	gap of unknown length
102912	106305:	contig of 3394 bp in length
106306	106405:	gap of unknown length
106406	111977:	contig of 5572 bp in length
111978	112077:	gap of unknown length
112079	116021:	contig of 3944 bp in length
116022	116121:	gap of unknown length
116122	118407:	contig of 2286 bp in length
118408	118507:	gap of unknown length
118508	121535:	contig of 3028 bp in length
121536	121635:	gap of unknown length
121636	125077:	contig of 3442 bp in length
125078	125177:	gap of unknown length
125178	129179:	contig of 4002 bp in length
129180	129279:	gap of unknown length
129280	132170:	contig of 2891 bp in length
132171	132270:	gap of unknown length

132271	134079:	contig of 1809 bp in length
134080	134179:	gap of unknown length
134180	137358:	contig of 3179 bp in length
137359	137458:	gap of unknown length
137459	139978:	contig of 2520 bp in length
139979	140078:	gap of unknown length
140079	142972:	contig of 2894 bp in length
142973	143072:	gap of unknown length
143073	145874:	contig of 2802 bp in length
145875	145974:	gap of unknown length
145975	147668:	contig of 1794 bp in length
147669	147668:	gap of unknown length
147670	149593:	contig of 1725 bp in length
149594	149693:	gap of unknown length
149694	151287:	contig of 1594 bp in length
151288	151387:	gap of unknown length
151388	153321:	contig of 1934 bp in length
153322	153421:	gap of unknown length
153422	155471:	contig of 2050 bp in length
155472	155571:	gap of unknown length
155572	157991:	contig of 2420 bp in length
157992	158091:	gap of unknown length
158092	160551:	contig of 2460 bp in length
160552	160651:	gap of unknown length
160652	162741:	contig of 2090 bp in length
162742	162841:	gap of unknown length
162842	164428:	contig of 1587 bp in length
164429	164528:	gap of unknown length
164529	166471:	contig of 1943 bp in length
166472	166571:	gap of unknown length
166572	168429:	contig of 1858 bp in length
168430	168529:	gap of unknown length
168530	170496:	contig of 1967 bp in length
170497	170596:	gap of unknown length
170597	172644:	contig of 2048 bp in length
172645	172744:	gap of unknown length
172745	174365:	contig of 1621 bp in length
174366	174465:	gap of unknown length
174466	175806:	contig of 1341 bp in length
175807	175906:	gap of unknown length
175907	178082:	contig of 2176 bp in length
178083	178182:	gap of unknown length
180094	180094:	contig of 1912 bp in length
180095	180194:	gap of unknown length
180195	181465:	contig of 1271 bp in length
181466	181565:	gap of unknown length
181566	183890:	contig of 2325 bp in length
183891	183990:	gap of unknown length
183991	185397:	contig of 1407 bp in length
185398	185497:	gap of unknown length
185498	187581:	contig of 2084 bp in length
187582	187681:	gap of unknown length
187682	188760:	contig of 1079 bp in length
188761	188860:	gap of unknown length
188861	190644:	contig of 1784 bp in length
190645	190744:	gap of unknown length
190745	191885:	contig of 1141 bp in length
191886	191985:	gap of unknown length
191986	193129:	contig of 1144 bp in length
193130	193229:	gap of unknown length
193230	194840:	contig of 1611 bp in length
194841	194940:	gap of unknown length
194941	196463:	contig of 1523 bp in length
196464	196563:	gap of unknown length

alignment_scores: Quality: 906.00 Length: 306
Ratio: 3.485
Percent Similarity: 84.967 Percent Identity: 52.941

alignment_block:
US-09-975-308-9 x AC096461

Align seg 1/1 to: AC096461 from: 1 to: 204143

```
2 AsnHisSerValValThrGluPheIleIleLeuGlyLeuThrIysIysPr 18
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
58436 AACGACACTGGTAAACAGAGTTCGCTTCAGGCGCTCTCTGAGCATCC 58485
18 ogIleuGlnGlyIleIlePheLeuPhePheIleValIleValIleVal 35
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
58486 GCAGTACACAGCTCTCTGTTTATCTGTTCTCTCTCTCTCTCTCTG 58535
35 IapHeuGlnGlyAsnMetLeuIleIleIleAlaIysIleTySerAsnThr 51
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
58536 CCTCTCAGGTATGTCCTTATCATCTTGGCCATTCACCTGCAACCTGG 58585
52 LeuHisThrProMetTyValPheLeuLeuThrLeuAlaIleValAspIle 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
58586 CTCACACACCCCAATGATTTTCTGTTCAATTTGGCTACATGATGT 58635
68 eIleCysThrThrSerIleIleProIysMetLeuGlyThrMetLeuThr 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
58636 CATCTGACTCTCTCCATTCATGCCAAAGGCCCTGAAGGCTCTGTGCA 58685
85 eIeGlnAsnThrIleSerTyValGlyCysMetSerGlnLeuPheLeuPhe 101
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
58686 AGAGAAACCTATCTCTATGCTGCTCATGACCATCCAGCTCTATTCT 58735
102 ThrTrpSerIleuGlyAlaGluMetValLeuPheThrThrMetAlaTyra 118
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
58736 ACATGGTCTGCTTCCTCAGAGTTCCTCCTCATGCTCATGGCTACGA 58785
118 PaqTyValAlaIleCysPheProLeuHisTySerThrIleMetAsnH 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
58786 CGCTTATGACGCATCTGCCATCTCTGCTATACACCATCATGACGA 58835
135 IShiMetCysValAlaLeuLeuSerMetValMetAlaIleAlaValThr 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
58836 AAGCATTTTGGCGGTGCTGCTGAGTGTGGCGCTTGTGCTTTC 58885
152 AsnSerTrpValHisThrAlaLeuIleMetArgLeuThrPheCysGlyPr 168
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
58886 AACACAGCCATTCACACAGAGTATGACGCGCTTGAAATTCGTGTGTC 58935
168 oAsnThrIleAspHisPhePheCysGlnIleProLeuLeuAlaLeu 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
58936 CAATGTCATTAGCAGCTTCTCTGTGAGTGCTCCCTGCTCTCTCT 58985
185 eIeCysSerProValArgIleAsnGluValMetValTyValAlaAspIle 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
58986 CCTGTAGCTCCACCTATGTGAAACAGTGTGATGTCTGTGCTGACGCC 59035
202 ThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyGlyPheIle 218
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59036 TTTTATGGCATTTGAACTTCCTGATGACCATCGTGTGATGCTTCAT 59085
218 eIleValAlaIleLeuArgIleArgThrValGluGlyIysArgIysAlaP 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59086 CATCTCCAGCATCTGACAGTGGGACTTCAGAAAGGAAGCAAAAGCCT 59135
235 heSerThrCysSerSerHisLeuThrValValThrLeuTyTySerPro 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59136 TCCTCTACCTGCTCTCCACCTCATCTGTGTGTCATGTATACACTGCT 59185
252 ValIleTyThrTyThrIleArgProAlaSerSerTyThrPheGluArgAs 268
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59186 GTCTTTATGCTTACATTAAGCCCTGCTCCAGCTATTAACGACAGAGAG 59235
268 pIysValValAlaAlaLeuTyThrLeuValThrProThrLeuAsnProM 285
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59236 CAAATTTGGCTGTGTCTGTATACATGCTGAGCCCTCAGCTCAACCTC 59285
285 eTyValTySerPheGlnAsnArgGluMetGlnAlaGlyIleArgIysVal 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59286 TGATCTACACTTGTAGAAACAAAGAGGTCAAGCAGCTCTCAGGAAATTT 59335
```

```
302 PheAlaPheLeuIysHis 307
||| |||:|||||:|||||:
59336 TTCCTTTCTCAGAAAT 59353
```


OM of: US-09-975-308-9 to: N_Geneseq_032802:* out_format : pfs

Date: Jun 24, 2002 12:59 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=framet+pn model -DEV=xlh  
-O=/cgn2_1/USFPO.spool/US09975308/runat_24062002_090304_26270/app_query.fasta.1.368  
-DB=N_Geneseq_032802 -QFMT=fastcap -SUFFX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsuum62  
-TRANS=human40.cdt -LIST=45 -DOCCALIGN=200 -THR_SCORE=pct  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USR=US09975308.@CGN1.1.0 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-975-308-9  
Query length: 307  
Database: N_Geneseq_032802:*  
Database sequences: 1736436  
Database length: 858457221  
Search time (sec): 209.350000
```

score list:

```
Sequence Strd Orig zScore EScore len 1 Documentation  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32037 + 1574.00 2913.00 6.6e-154  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542353 + 1574.00 2912.96 6.6e-154  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AA517179 + 1339.00 2476.78 1.3e-129  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542284 + 861.00 1584.17 6.8e-80  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA52952 + 861.00 1581.75 9.3e-80  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH31706 + 860.00 1583.00 7.9e-80  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542455 + 731.00 1342.38 2.0e-66  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542395 + 725.00 1332.76 2.5e-65  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH61132 + 725.00 1331.47 8.1e-66  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32092 + 723.00 1327.44 1.4e-65  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32067 + 707.00 1297.98 6.0e-64  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542407 + 707.00 1296.63 7.1e-64  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32033 + 706.00 1296.12 7.6e-64  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542313 + 699.00 1282.84 4.2e-63  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542390 + 699.00 1282.54 4.3e-63  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH31761 + 697.00 1282.80 4.2e-63  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH31886 + 695.00 1275.52 1.1e-62  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542439 + 693.50 1272.40 1.6e-62  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32097 + 693.00 1271.76 1.7e-62  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH31177 + 692.00 1273.48 1.4e-62  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32044 + 691.00 1268.06 2.8e-62  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH31761 + 691.00 1267.97 2.8e-62  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH33815 + 691.00 1214.83 2.6e-59  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32035 + 687.00 1260.22 7.6e-62  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542350 + 687.00 1259.41 8.4e-62  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH19150 + 686.50 1259.42 8.4e-62  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32144 + 686.50 1258.93 8.9e-62  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH13717 + 686.00 1262.31 5.8e-62  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH31187 + 686.00 1258.59 9.3e-62  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32010 + 686.00 1258.59 9.3e-62  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542357 + 685.00 1253.02 1.9e-61  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH78085 + 685.00 1253.02 1.9e-61  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32440 + 683.50 1254.21 1.6e-61  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32387 + 683.50 1253.94 1.7e-61  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AAH16617 + 683.50 1253.51 1.8e-61  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAH31037 + 683.50 1251.40 2.3e-61  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH44109 + 683.50 1246.17 4.6e-61  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH54562 + 683.50 1246.17 4.6e-61  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH24345 + 683.50 1246.17 4.6e-61
```

```
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAK02855 + 683.50 1246.17 4.6e-61  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAK82829 + 683.50 1246.17 4.6e-61  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH12861 + 683.50 1246.17 4.6e-61  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH34218 + 683.50 1246.17 4.6e-61  
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH27777 + 683.50 1246.17 4.6e-61  
seq_name: /SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32037
```

seq_documentation_block:

```
ID AAH32037 standard; DNA; 921 BP.  
XX  
XX AAH32037;  
AC  
XX  
XX  
DT 30-JUL-2001 (first entry)  
XX  
DE Human olfactory receptor polynucleotide, SEQ ID NO: 610.  
XX  
XX Human; olfactory receptor; OR; primary scent determination;  
KW secondary scent determination; polypeptide library; odour receptor;  
KW scent profile; scent fingerprint; scent representation; de.  
XX  
XX Homo sapiens.  
OS  
XX  
PN WO200127158-A2.  
XX  
XX 19-APR-2001.  
PD  
XX  
PF 06-OCT-2000; 2000MO-US27582.  
XX  
XX 08-OCT-1999; 99US-0158615.  
PR 24-FEB-2000; 2000US-0184809.  
XX  
XX (DIGT-) DIGSCENTS.  
PA (YEDA ) YEDA RES & DEV CO LTD.  
XX  
XX Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;  
PI WPI; 2001-290713/30.  
XX  
XX New polynucleotides which encode polypeptides involved in olfactory  
PT sensation for identifying olfactory agonists and antagonists -  
XX  
XX  
XX Claim 8; Page 425; 1857pp; English.
```

The present sequence is one of a number of isolated polynucleotides which encode polypeptides involved in olfactory sensation. The polynucleotides can be used in screening for olfactory agonists and antagonists. The methods allow for the determination of primary scents and the identification of the odour receptors used to detect these primary scents. The methods also enable determination of secondary scents and the identification of combinations of odour receptors that are involved in detecting such secondary scents. This enables the construction of a scent representation (also called a scent fingerprint or scent profile), which may be used to re-create and edit scents. Libraries of olfactory receptors are useful for determining the interaction pattern of a composition with the receptors, and can be used for determining differences in the olfactory faculties of different individuals.

Sequence 921 BP; 218 A; 235 C; 183 G; 285 T; 0 other;

alignment_scores:

```
Quality: 1574.00 Length: 307  
Ratio: 5.127 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.349
```

alignment_block:

```
US-09-975-308-9 x AAH32037 ..
```

```
Align seg 1/1 to: AAH32037 from: 1 to: 921
```

```
1 MetashtiservValtrhGluphelelleucllyLeurhlyslly 17
```

```

|||||
1 ATGAATCCAGGCGTGTAACTGAGTTCATTATTCGGGCTCAGCAAAA 50
17 sprogileuGinglyIleIlePheLeuPhePheIleValTyrLeu 34
|||||
51 GCCTGAACGCCAGGAATTATCTCTCTTTTCTCATGTGCTATCTTG 100
34 aIaIaPheLeuGlyAsnMetLeuIleIleIleAlaIysIleTyrSerAsn 50
|||||
101 TGGCTTTTCTGGCAACATGCTCATCTCATCTTCCAAATCTATACAC 150
51 ThrLeuHisThrProMetTyrValPheLeuLeuThrLeuAlaValAs 67
|||||
151 ACCTTGATACGCCCATGTATGTTTCTCTGACACTGGCTGTGGGA 200
67 pIleIleCysThrThrSerIleIleProIysMetLeuGlyThrMetLeu 84
|||||
201 CATCATGTGCACACACAGACATCATCCAGAGATGCTGGGAGCATGCTAA 250
84 hrSerGluAsnThrIleSerTyrAlaGlyCysMetSerGlnLeuPheLeu 100
|||||
251 CATGAGAAATACCATTTCAATATCAGAGCTGACATGCCAGCTTCTTG 300
101 PheThrTrpSerLeuGlyAlaGluMetValLeuPheThrThrMetAlaTyr 117
|||||
301 TTCACATGTCTCTGGAGCTGAGATGGTCTCTTCCACCAACATGGCCTA 350
117 rAspArgTyrValAlaIleCysPheProLeuHisTyrSerThrIleMetAla 134
|||||
351 TGACCGCATGTGGCCATTTGTTCCCTTCATTACAGTACTGTATGA 400
134 snHisHisMetCysValAlaLeuLeuSerMetValMetAlaIleAlaVal 150
|||||
401 ACCACCATATGTGTAGCTTGTGACATGTCACATGTCATGGCTATTCAGCTC 450
151 ThrAsnSerTrpValHisThrAlaLeuIleMetArgLeuThrPheCysGln 167
|||||
451 ACCAATCTCTGGGTGCACACAGCTTATCATAGAGTGTGACTTCTGTGG 500
167 yProAsnThrIleAspHisPhePheCysGluIleProProLeuAlaIle 184
|||||
501 GCCAAACACCATTTGACACTTCTGTGTAGTACCCCATTTGCTGGCTT 550
184 euSerCysSerProValArgIleAsnGluValMetValTyrValAlaAsp 200
|||||
551 TGTCTGTGACCCCTGTAGAAATCAATGAGGTGATGTGTCTGTAT 600
201 IleThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyrGlyPhe 217
|||||
601 ATTACCTGTGGCATAGGGAGCTTTTATCTTACCTGATCTCTATGTGTT 650
217 eIleIleValAlaIleLeuArgIleArgThrValGluGlyAsArgIysA 234
|||||
651 TATCATTTGTTGCTATCTCCGTATCCGACAGTAGAAGCAAGAGAAAG 700
234 IaPheSerThrCysSerSerHisLeuThrValValThrLeuTyrTyrSer 250
|||||
701 CCTCTCAACATGCTCATCTCATCTCAGAGTGTGACCCCTTACTATTTCT 750
251 ProValIleIleThrTyrIleArgProAlaSerSerTyrThrPheGluArg 267
|||||
751 CCGTATATCTACACCTATATCCGCTGCTCCAGCTATATCATTTGAAG 800
267 gAspIysValValAlaIleLeuTyrThrLeuValThrProThrLeuAsn 284
|||||
801 AGACAGAGGTGTAGCTGCACCTATATCTTGTGACTCCACATTTAAACC 850
284 rOmetValTyrSerPheGlnAsnArgGluMetGlnAlaGlyIleArgIys 300
|||||
851 CGATGTGTGTACAGCTTCCAGAAATAGGAGATGCAAGGAGAAATTAGGAG 900
301 ValPheAlaPheLeuIysHis 307
|||||

```

901 GTGTTGCATTTCGAACAC 921

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542323

seq_documentation_block:

ID AA542323 standard; cDNA: 924 BP.

XX AA542323;

XX 18-DEC-2001 (first entry)

XX Human cDNA encoding olfactory receptor AOR124.

XX Human: olfactory receptor; G protein-coupled receptor; GPCR; odourant; ss; food additive; cosmetic; fragrance; pharmaceutical additive.

XX Homo sapiens.

XX WO200168805-A2.

XX 20-SEP-2001.

XX 13-MAR-2001; 2001WO-US07771.

XX 13-MAR-2000; 2000US-0188914.

XX 24-MAR-2000; 2000US-0192033.

XX 12-APR-2000; 2000US-0198474.

XX 24-APR-2000; 2000US-0199335.

XX 26-MAY-2000; 2000US-0207702.

XX 23-JUN-2000; 2000US-0213849.

XX 16-AUG-2000; 2000US-0228534.

XX 07-SEP-2000; 2000US-0230732.

XX 07-FEB-2001; 2001US-0266862.

XX (SENO-) SENOMYX INC.

XX Zozulya S;

XX WPI: 2001-570867/64.

XX P-PDB: AAU24630.

XX Nucleic acids encoding human olfactory G protein-coupled receptors, useful for screening for compounds involved in olfactory sensation, where the compounds can be used in the food, pharmaceutical and cosmetic industries to customise odours.

XX Claim 1; Page 135; 319pp; English.

XX The invention relates to nucleic acids encoding human olfactory receptors, OR, (a G protein-coupled receptor, GPCR). The OR's specifically recognise molecules, odourants, that elicit specific olfactory sensation. The human olfactory receptors and polynucleotides encoding them are useful for screening a library of chemical compounds for compounds that are involved in olfactory sensation. Modulators of their activity are useful for pharmacological and genetic modulation of olfactory signalling pathways. Therefore, they can be used in the food, pharmaceutical and cosmetic industries to customise odours and fragrances. The present sequence encodes a human olfactory receptor of the invention.

XX Sequence 924 BP; 219 A; 235 C; 184 G; 286 T; 0 other;

alignment_scores:

Quality: 1574.00	Length: 307
Ratio: 5.127	Gaps: 0
Percent Similarity: 100.000	Percent Identity: 99.349

alignment_block:

US-09-975-308-9 x AA542323 ..

Align seg 1/1 to: AA542323 from: 1 to: 924


```

1 MetAsnHisSerValValThrGluPheIleIleLeuGlyLeuThrLysLys 17
1 ATGAATCAGAGCGCTTGAAGTTCATTTATCTGGGCTCACCAAAAA 50
17 sProGluLeuGlyIleIlePheLeuPhePheLeuIleValTyrLeu 34
51 GCCTGAAGCTCAGGAATATCTCCCTTTTTCATCTGCTATCTG 100
34 aAlaPheLeuGlyAsnMetLeuIleIleAlaIleValTyrSerAsn 50
101 TGGCTTTTCTGGCAACATGCTCATCATTCGCAAAATCTATACAC 150
51 ThrLeuHisThrProMetTyrValIlePheLeuThrLeuAlaValAs 67
151 ACCTTCATACGCCCATGTAAGTTTCTTCACACTGGCTGTGTGGA 200
67 PileIleCysThrThrSerIleIleProLysMetLeuGlyTyrMetLeu 84
201 CATCATCTGCACAAACAGATCATACGAAAGTGGGGACCATGTAA 250
84 hSerLysAsnThrIleSerTyrAlaGlyCysMetSerGlnLeuPheLeu 100
251 CATCAAAAATACCATTTTCATATGACGGCTGATGCCAGCTCTCTTG 300
101 PheThrTyrSerLeuGlyValGluMetValLeuPheThrThrMetAla 117
301 TTCACATGCTCTGGAGCTGAGATGTTCTTCACACCATGGCTTA 350
117 rAspArgTyrValAlaIleCysPheProLeuHisTyrSerThrIleMet 134
351 TGACCGCTATGAGCCATTGTTCCCTCTTCATACAGTACGTTATGA 400
134 snHisHisMetCysValAlaLeuLeuSerMetValMetAlaIleAla 150
401 ACCACCATATGTTGTAGCTTGTCTCAGCATGGCTCATGGCTATTCG 450
151 ThrAsnSerTyrValHisThrAlaLeuIleMetArgLeuThrPheCys 167
451 ACCAATTCCTGGGTGCACACAGCTTTCATCATGAGTTGACTTCTGTG 500
167 yProAsnThrIleAspHisPhePheCysGluIleProProLeuLeuAl 184
501 GCCAAACACCATGACCACTTCTGTGATGATACCCCATCTGCTGCTT 550
184 eSerCysSerProValArgIleAsnGluValMetValTyrValAlaAsp 200
551 TGTCTGTAGCCCTGTAAGAATCAATGAGGTGATGTGTGTGTGAT 600
201 IleThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyrGly 217
601 ATTACCCCTGGCATAGGGACTTATCTTACCTGCATCTCCATATGTTT 650
217 eIleIleValAlaIleLeuArgIleArgThrValGluGlyLysArgLys 234
651 TATCATTTGTTGCTATCTCCGTAATCCGACAGTAGAAGCAAGAGAAG 700
234 laphSerThrCysSerSerHisIleuThrValValThrLeuTyrTyrSer 250
701 CCTTCACACATGCTCATCTCATCTCAGTGTGACCTTCATCTATCTT 750
251 ProValIleTyrThrTyrIleArgProAlaSerSerTyrThrPheGlu 267
751 CCGTATATCTACACCTATATCGCCCTGCTTCAGCTATATACATTTAA 800
267 gAspLysValValAlaIleLeuTyrThrLeuValThrProThrLeuAsn 284
801 AGACAAAGGTGAGTGTGACTCTATCTGTGACTCCACATTTAAACC 850
284 rOMeValTyrSerPheGlnAsnArgGluMetGlnAlaGlyIleArgLys 300
851 CGATGCTGTACAGCTTCAGAAATGAGGAGATGAGGAGAAATTTAGGA 900
301 ValPheAlaPheLeuLysHis 307

```

```

|||||
901 GTGTTTCATTTCTGAAACAC 921
seq_name: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT:AA517179
seq_documentation_block:
ID AA517179 standard; cDNA; 798 BP.
XX
AC AA517179;
XX
DE 14-FEB-2002 (first entry)
XX
Human cDNA encoding a novel G protein-coupled receptor, NOV6.
XX
KW Human; ss; G protein-coupled receptor; GPCR; NOV6; cardiant;
KW antiarteriosclerosis; antibacterial; virucide; fungicide; protozoacide;
KW antidiabetic; cytotoxic; analgesic; antisthmatic; nootropic;
KW neuroprotective; antiinflammatory; gene therapy; transgenic animal;
KW cardiomyopathy; atherosclerosis; infection; pain; anorexia; bulimia;
KW asthma; neurological disorder; Parkinson's disease; stroke;
KW Alzheimer's disease; multiple sclerosis; lesionial psoriatic skin;
KW ischaemia; cirrhotic hepatitis; acute pancreatitis; diabetes; cancer;
KW angiogenesis; obesity; olfactory disorder; chromosome 11.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 5'UTR 1..23
FT /*tag= a
FT CDS 24..782
FT /*tag= b
FT /*product= "NOV6"
FT 3'UTR 783..798
FT /*tag= c
XX
WO20017177-A2.
XX
PD 18-OCT-2001.
XX
PF 11-APR-2001; 2001WO-US11901.
XX
PR 11-APR-2000; 2000US-195994P.
PR 11-APR-2000; 2000US-196538P.
PR 26-APR-2000; 2000US-199902P.
PR 26-APR-2000; 2000US-199964P.
PR 27-APR-2000; 2000US-199948P.
PR 27-APR-2000; 2000US-199956P.
PR 27-APR-2000; 2000US-200176P.
PR 17-JUL-2000; 2000US-218995P.
PR 25-JUL-2000; 2000US-220644P.
PR 04-JAN-2001; 2001US-259641P.
PR 29-JAN-2001; 2001US-264851P.
PR 14-FEB-2001; 2001US-268567P.
PR 13-MAR-2001; 2001US-0220644.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shenoy S, Gangoli EAG, Rastelli L, Smithson G, Padigaru M;
PI Vernet CAM, Wolenc AR, Casman SJ, Tchernev VT, Szekeres ES;
PI Gorse W, Alsobrook JP, Burgess CE;
XX
DR MPI: 2002-041291/05.
DR P-PSDB: AAU11099.
XX
PT Human G-protein coupled receptors, NOV1-12, useful for diagnosis and
PT treatment of e.g. cardiomyopathy, also in screening for specific
PT modulators -
XX
PS Claim 11; Page 30; 171pp; English.
XX
CC The invention relates to Human G-protein coupled receptor (GPCR)
CC polypeptides, designated NOV1-12, nucleic acids encoding them,

```


CC their activity are useful for pharmacological and genetic modulation of
CC olfactory signalling pathways. Therefore, they can be used in the food,
CC pharmaceutical and cosmetic industries to customise odours and
CC fragrances. The present sequence encodes a human olfactory receptor of
CC the invention.

Sequence 987 BP; 190 A; 309 C; 235 G; 253 T; 0 other;

alignment_scores:

Quality:	861.00	Length:	306
Ratio:	3.337	Gaps:	0
Percent Similarity:	84.314	Percent Identity:	49.673

alignment_block:

US-09-975-308-9 X AAS42284 . .

Align seg 1/1 to: AAS42284 from: 1 to: 987

[illegible]

```

717  CGCTCCAGCATCTCTGAAGGTGAAAGCATCTGCGCTGGGGAGGACGAAAGCCT 766
235  heserThrCysSerSerHisLeuThrValValThrLeuTyTySerPro 251
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
767  TCTCCACACTCTCTCTCCACCTCCACCGCTGGGTGCTGATTTACACCGCT 816
252  ValIleTyThrTyTleArgProIleSerSerTyThrPheGluArgAs 268
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
817  GTCTTCTACGCCCTACATAAGCCCGCTCTCTGGCTACAGCGCAGGGAAG 866
268  pIysValValAlaIleLeuTyThrLeuValThrProThrLeuAsnProM 285
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
867  CAAGTGGCGCGCCCTGCTGTACACGTGCTGAGTCTTACCTCCACACCCC 916
285  eValTyTyrSerPheGlnAsnArgIleMetGlnIleAlaGlyLeuArgVal 301
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
917  TCAATCTACTTTTGAAGAACAAGAGGTCTCAAAAGCAGCCCTCAGAAAGCTT 966
302  PheAlaPheLeuLysHis 307
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
967  TTCCCTTTCTTCAGAAAT 984
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:ADD1

```

seq_documentation_block:

ID AAD12952 standard; cDNA; 1241 BP.

AC AAD12952;

DT 16-OCT-2001 (first entry)

DE Human G-protein coupled receptor-9 (GCREC-9) cDNA.

KM Human; protein coupled receptor-9; GPCR-9; glycolytic; hepatotropic;
 KM virulence; antiinflammatory; anticonvulsant; antileptic; neuroprotective;
 KM nocutopic; cerebroprotective; hypotensive; tranquilliser; vulnerary;
 KM ophthalmologic; cell proliferative disorder; actinic keratosis;
 KM anorectic; arteriosclerosis; atherosclerosis; cirrhosis; hepatitis;
 KM psoriasis; cancer; neurological disorder; stroke; Alzheimer's disease;
 KM Huntington's disease; Parkinson's disease; cardiovascular disorder;
 KM epilepsy; hypertension; varicose vein; vasculitis; dysphagia; dyspepsia;
 KM anorexia; gastrointestinal disorder; pancreatitis; autoimmune disorder;
 KM Addison's disease; Crohn's disease; acquired immune deficiency syndrome;
 KM AIDS; uvelitis; infection; trauma; metabolic disorder; diabetes; obesity;
 KM osteoporosis; transgenic animal; gene therapy; ss.

05 Homo sapiens.

FH	Key	Location/qualifiers
----	-----	---------------------

FT. / * t a q = a

XX
F1
/product- human secretory protein

PN WO200157085-A2

PD 09-AUG-2001.

PF 01-FEB-2001; 2001WO-US03455.

02-FEB-2000: 2000US-0180093.

11-FEB-2000; 200005-0182043-
PK
XX

PA (INCY-) INCYTE GENOMICS INC.

PI Baughn MR, Au-Young J, Yue H,

DR WPI; 2001-488869/53.

XX

PT preventing and treating

PT disorders - cardiovascular, gastro

PT disorders -

XX WPI: 2001-290713/30.

PT New polynucleotides which encode polypeptides involved in olfactory
PT sensation for identifying olfactory agonists and antagonists -

PS Claim 8; Page 289; 1857pp; English.

XX The present sequence is one of a number of isolated polynucleotides
CC which encode polypeptides involved in olfactory sensation. The
CC polynucleotides can be used in screening for olfactory agonists and
CC antagonists. The methods allow for the determination of primary
CC scents and the identification of the odour receptors used to detect
CC these primary scents. The methods also enable determination of
CC secondary scents and the identification of combinations of odour
CC receptors that are involved in detecting such secondary scents.
CC This enables the construction of a scent representation (also called
CC a scent fingerprint or scent profile), which may be used to re-create
CC and edit scents. Libraries of olfactory receptors are useful for
CC determining the interaction pattern of a composition with the receptors,
CC and can be used for determining differences in the olfactory faculties
CC of different individuals.

XX Sequence 924 BP; 170 A; 293 C; 219 G; 242 T; 0 other;

alignment_scores:

Quality: 860.00 Length: 305
Ratio: 3.346 Gaps: 0
Percent Similarity: 84.262 Percent Identity: 49.836

alignment_block:
US-09-975-308-9 x AAH31706 ..

Align seg 1/1 to: AAH31706 from: 1 to: 924

2 AsnHisSerValValThrGluPheIleIleLeuGlyLeuThrLysLysSpr 18
10 AACCAAGACGTTGGTAAACGAGTTCACGAGCGCTTTGGAGACCC 59
18 OGluLeuGlnGlyIleIlePheLeuPhePheLeuIleValTyrLeuValA 35
60 AGAAATAACCGGGGTCTCTTATTCAGCTGTTCCTCTCTCTACTCTGGG 109
35 lApheLeuGlyAsnMetLeuIleIleIleAlaLysIleTyrSerAsnThr 51
110 CCTCTACAGGTAATGCTCTCATCCTTGCGCATTCACGTTCAACCTGGG 159
52 LeuHisThrProMetTyrValPheLeuLeuThrLeuAlaValAlaSpr 68
160 CTCACAGCGCTCATGATGTTTCTTACCACTTGCTGCTACTATGACAT 209
68 eIleCysThrThrSerIleIleProLysMetLeuGlyThrMetLeuThrS 85
210 TATCTGCACCTCTCTCATCATGCCCAAGCGCTGGCCAGCTGGTGG 259
85 eArgLysAsnThrIleSerTyrAlaGlyCysMetSerGlnLeuPheLeu 101
260 AAGAGAGCTCCATCTCTACGGGGGCTGATGGCCAGCTCTATTCCTC 309
102 ThrTrpSerLeuGlyAlaGluMetValLeuPheThrThrMetAlaTyrAs 118
310 ACGTGGCGCATCCCTACAGCTGCTGCTCTACAGCTGATGCGCATAT 359
118 PArgTyrValAlaIleCysPheProLeuHisTyrSerThrIleMetAsn 135
360 CCGGACGAGGACATCTGCCCGCTGCATTAACACACATGATGAGA 409
135 lAsnMetCysValAlaLeuLeuSerMetValMetAlaIleAlaValThr 151
410 AGGTGTTCTGACAGCGGGCTGGCCACGCGGTGGCTGCTGGCGGTC 459
152 AsnSerTrpValHisThrAlaLeuIleMetArgLeuThrPheCysGlyPr 168

460 AACACGGCCATCCACACGGGGCTATGCTGCGCTGGATTTCTGCGCC 509
168 oAsnThrIleAspHisPhePheCysGluIleProProLeuLeuAlaLeuS 185
510 CAATGCTATATCCATTTCTTCTGAGGTCGCCCTGCTGCTCTCT 559
185 eCysSerProValArgIleAsnGluValMetValTyrValAlaAspIle 201
560 CCTGCACCTCCATCCATACGCTGTCATGATGTCGCGGATGCT 609
202 ThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyrGlyPheI 218
610 TTCACGGCATATGAGACTTCTCATATACATGCCGCTCATGCTTCAT 659
218 eIleValAlaIleLeuArgIleArgThrValGluGlyLysArgLysAlaP 235
660 CGTCTCAGCATCTCTGAAGTGAAGACTGCTGGGGAGGCAAAACCT 709
235 hEserThrCysSerSerHisLeuThrValAlaThrLeuTyrTyrSerPro 251
710 TCTCCACCTGCTCTCCACCTCACCGTGGTGCATGATATACACGCT 759
252 ValIleTyrThrTyrIleArgProAlaSerSerTyrThrPheGluArgAs 268
760 GTCTTACAGCCTATACATTAAGCCGCTCTGCGCTACAGCGAGGAGAG 809
268 pLysValValAlaAlaLeuTyrThrLeuValThrProThrLeuAsnPro 285
810 CAAGTGGCGCGCTGCTGTACACTGTGCTGAGCTCACCTCAACCCCC 859
285 eValTyrSerPheGlnAsnArgLysGluMetGlnAlaGlyIleArgLysVal 301
860 TCATCTTACTTTCAGAAACAGAGAGGTCAAGACGCTTCAGAGACTT 909
302 PheAlaPheLeuLys 306
910 TTCCTTCTTCAGA 924
seq_name: /STD1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.AAS42455
seq_documentation_block:
ID AAS42455 standard; cDNA; 954 BP.
AC AAS42455;
DT 18-DEC-2001 (first entry)
XX Human cDNA encoding olfactory receptor AOLFRL30B.
DE Human; olfactory receptor; G protein-coupled receptor; GPCR; odourant;
KW ss; food additive; cosmetic; fragrance; pharmaceutical additive.
OS Homo sapiens.
PN WO200168805-A2.
PD 20-SEP-2001.
XX 13-MAR-2001; 2001WO-US07771.
PF 13-MAR-2000; 2000US-0188914.
PR 24-MAR-2000; 2000US-0192033.
PR 12-APR-2000; 2000US-0198474.
PR 24-APR-2000; 2000US-0199335.
PR 26-MAY-2000; 2000US-0207702.
PR 23-JUN-2000; 2000US-0213849.
PR 16-AUG-2000; 2000US-0226534.
PR 07-SEP-2000; 2000US-0230732.
PR 07-FEB-2001; 2001US-026862.
XX (SENO-) SENOMYX INC.
XX PA

xx xx zozulya S:
DR MPI: 2001-570867/64.
P-PSDB; AAU24762.

xx
xx Nucleic acids encoding human olfactory G protein-coupled receptors,
PT useful for screening for compounds involved in olfactory sensation,
PT where the compounds can be used in the food, pharmaceutical and
PT cosmetic industries to customise odours -
xx
xx Claim 1; Page 194; 319pp; English.

xx The invention relates to nucleic acids encoding human olfactory
CC receptors, OR, (a G protein-coupled receptor, GPCR). The OR's
CC specifically recognise molecules, odorants, that elicit specific
CC olfactory sensation. The human olfactory receptors and polynucleotides
CC encoding them are useful for screening a library of chemical compounds
CC for compounds that are involved in olfactory sensation. Modulators of
CC their activity are useful for pharmacological and genetic modulation of
CC olfactory signalling pathways. Therefore, they can be used in the food,
CC pharmaceutical and cosmetic industries to customise odours and
CC fragrances. The present sequence encodes a human olfactory receptor of
CC the invention.

xx Sequence 954 BP; 224 A; 241 C; 182 G; 307 T; 0 other;
SQ

alignment_scores: Quality: 731.00 Length: 300
Ratio: 3.137 Gaps: 0
Percent Similarity: 77.667 Percent Identity: 45.000

alignment_block:
US-09-975-308-g x AAS42455 ..

Align seg 1/1 to: AAS42455 from: 1 to: 954

2 AsnHisSerValThrGluPheIleLeuGlyLeuThrLysLysPr 18
 |||||..... :|||.....
13 AATCAACAGCTCATTCGAATCATCATCTTGGGGTGACCACCTGAA 62
18 cgluenglnclyllellepheleuphepheleulevalylreueVala 35
 |||||..... :|||.....
63 TGAATTGCGAATTACTCTTCACCATCTTCTTTGTGCCTACATATGCA 112
35 laphelenglyAsmMetleullellelealalySileTySerAnthr 51
 ::::: |||||..... :|||.....
113 CTTNAGAGGCCAATGTTTTTATCATGTGGGACCATGATGATCCCAC 162
52 leuhisthrPromettyValPheleuleuthrIleualValaVasPl 68
 |||||..... :|||.....
163 CTACACACACCATGTAATTTCTCCTAGAAATCTTGCCCTTATTGACAT 212
68 ellecysthrThrserrilleleProLysMetleuglyThrMetLeuthr 85
 |..... :|||..... :|||.....
213 CTGCTACACTACTACTAATGTCCCCACAATGATGGGCAATCTTGCTGAG 262
85 ergLuasnThrIlesertyAlaglyCySmetsErGlneupheleupe 101
 :::..... :|||..... :|||.....
263 AGAGAAATCATTTCTCTATGGAGGCTGTGTACCCACAGCTCTTTGCATTGC 312
102 ThrTPserLeuGlyAlaGluMetValleupheThrThrMetAlaTyraS 118
 ::: :|||..... :|||.....
313 ATTTCTTTTGGTGGCTCAGAAGTGTCTCCCTCCGCGACAAATGGCAATAGA 362
118 pArtyrValAlaIalecysPheProLeuHisTySerThrIleMetAsnH 135
 |||||..... :|||.....
363 TGCAATATATGCTATCTGAAGCCGTTAAGTACTATTTATATATAACA 412
135 IshisMetcySValAlaleuleusetMetValmetAlalleleAlaValThr 151
 :::..... :|||.....
413 AGGCGCTGTGACGCTGGTAGCACGCTCATGCTGGACATGTGGGTTTCTC 462

152	AsnSerTrpValHisThrAlaLeuIleMetArgLeuThrPheCysGlyPro	168
153	: : : :	
163	AsnCGAGTGGTGACACCCGCTTGTGACCTTGACCGCCCTTGCTGGTAA	512
164	: : : :	
168	oasnThrIleAspHisPhePheCysGluIleProProLeuLeuAlaLeuS	195
169	: : : :	
513	CAATCAGATCAATATTCTTCTTGAGATCACTCCCTGCTCATCTTGT	562
185	ercSerProValArgIleAsnGluValMetValTyrValAlaAspIle	201
186	: : : : :	
563	CTTGCTGATGATCTCCCTCCCAATGAACAGCGCTTGCTGCATTGGAGCT	612
202	ThrIleuAlaIleGlyAspPheIleuThrCysIleSerTyrGlyPheI	218
613	CTCATTAAGCTGGACTCTTCTTCCGTGCAATCAATCCTTCTCATCTTACAT	662
218	elIleValAlaIleLeuArgIleArgThrValGluGlyIleArgGlyAlaP	235
663	: : : :	
235	heserThrCysSerSerHisIleuThrValValThrIleuTyrTyrSerPro	251
713	TTTTCCACCTGGCTCCACCTCCATCTGCTATGTATTCTCATTAAGGCAGT	762
252	ValIleTyrThrTyrIleArgProAlaSerSerTyrThrPheGluArgAs	268
763	GCTATCTTCAAGATGTGAGGCCCATCATCTATCTCTGAGAAAGA	812
268	PlysValValAlaAlaLeuTyrThrIleuValThrProThrIleuAsnProM	285
813	TAGATTGATCTCAGTCTGTATAGTGTGTGCACACCCATGCTGATCTCTG	862
285	etValTyrSerPheGluAsnArgGluMetGlnAlaGlyIleArgIleVal	301
863	TAAATTATAGCGCTAAGGATATAGACATCAAGAAGCGCTGTGAAGGCCATA	912
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF61132		
seq_documentation_block:		
ID	AAF61132 standard; cDNA; 3600 BP.	
AC	AAF61132:	
XX		
XX	17-MAY-2001 (first entry)	
DT		
DE	Human OLFXY cDNA.	
XX		
KW	OLFXY; gene therapy; olfactory receptor; human; ds.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1179..2144
FT		/*tag= a
FT		/product= "OLFXY"
XX		
PN	DE19937839-A1.	
XX		
PD	15-FEB-2001.	
XX		
PF	13-AUG-1999; 99DE-1037839.	
XX		
PR	13-AUG-1999; 99DE-1037839.	
XX		
PA	(BRUE/) BRUES M.	
XX	(BOEN/) BOENISCH H.	
PI	Brues M, Boenisch H;	
XX		
DR	WPI: 2001-184006/19.	
DR	P-PSDB: AAB46999.	
PT	New human gene OLFXY, encoding an olfactory receptor, useful for	

PT diagnosis, treatment and development of pharmaceuticals -
 XX
 PS Disclosure; Page 3; 4pp; German.
 XX
 CC This invention describes a novel human OLFYX gene (I) and its encoding
 CC protein. The invention also describes (1) transcription factors, RNA
 CC polymerases, pharmaceuticals and chemicals that modulate (I), positively
 CC or negatively; (2) mRNA, and its splice variants or isoforms, transcribed
 CC from (I); (3) cDNA derived from the mRNA of (2) or from intron-less
 CC genes; (4) protein (II) derived from the mRNA, (I) or cDNA; (5)
 CC monoclonal antibodies or antisera directed against the whole of (II) or
 CC or more of its epitopes; (6) expression systems that produce native
 CC or recombinant (II); (7) ligand-binding studies or screening assays using
 CC native or recombinant (II) or cells, or cell membranes, that contain
 CC (II); (8) transgenic and knockout animals that express (II) at altered
 CC levels or not at all; (9) gene therapy methods which use (II), (I) or its
 CC related mRNA or DNA; (10) (anti)sense oligonucleotides derived from (I);
 CC and (11) diagnosis and treatment of diseases in which (II) is directly,
 CC or indirectly, involved. (I) encodes an olfactory receptor. (I), and its
 CC derived proteins, mRNA, cDNA and antisense, or sense, oligonucleotides
 CC are useful for treatment and diagnosis of diseases in which (I) is
 CC directly, or indirectly, involved, and for development and evaluation of
 CC new pharmaceuticals or technologies.
 CC
 XX
 SQ Sequence 3600 BP; 1119 A; 655 C; 588 G; 1238 T; 0 other;

alignment_scores:
 Quality: 728.00 Length: 300
 Ratio: 3.085 Gaps: 0
 Percent Similarity: 78.667 Percent Identity: 45.000

alignment_block:
 US-09-975-308-9 x AAF61132 ..

Align seg 1/1 to: AAF61132 from: 1 to: 3600

2 AsnHISserValValThrcGluPheIleIleLeuGlyLeuThrlsLysPr 18
 1191 AATCAACAGCTATATCTGAATTCATCATCTGGGATTCCTCAACCTTAA 1240
 18 cGluLeuGlnGlyIleIlePheLeuPhePheLeuValValValVal 35
 1241 TGAATTGCAAGTTTACTTATTCACCATCTCTTCTGCACTTATTTCTGTA 1290
 35 lApheLeuGlnAsnMetLeuIleIleIleAlaLysIleTyrSerAsnThr 51
 1291 CTTTGGAGGAAATATATTAATTATCTTGACAGCTGTGACTGATCCACAC 1340
 52 LeuHISThrPrometTyrValPheLeuLeuThrLeuAlaValAlaAsp11 68
 1341 CGGCAATACCCATATATATATTTCTAGGAACTGGCTTATTTGACAT 1390
 68 eIleCyStrThrSerIleIlePheProLysMetLeuGlyThrMetLeuThrs 85
 1391 CTGCTACACACACACATATGTCCTCCAGATGATGTCACCTCTCTCA 1440
 85 eRGLAsnThrIleSerTyrAlaGlyCysMetSerGlnLeuPheLeuPhe 101
 1441 AGAAAAAAGCATTTCTTATGTGGGGTGTGTGTCACACTTTTGCATTT 1490
 102 ThrTrpSerLeuGlyAlaGluMetValLeuPheThrThrMetAlaTyrAs 118
 1491 GTTTCCTTGTGGATCAGAGTGTCTCTACAGCAGCATATGACATATGA 1540
 118 PArGtyrValAlaIleCysPheProLeuHISThrIleMetAsnH 135
 1541 TCGTACATTCGCAATCTGCAATCCTTAAAGTATTCAGTTATTCGAGCA 1590
 135 lSHISMetCysValAlaLeuLeuSerMetValMetAlaIleAlaValThr 151
 1591 AGGTTCTATGCATATATAGCAGCCTCATGCTGGCTGCTGTTCCCTT 1640

152 AsnSerTrpValHISThrAlaLeuIleMetArgLeuThrPheCysGlyPr 168
 1641 AACTCAGTGTGCATACAGTGTGACATTTGCTGCTGCTTCTGTGCA 1690
 168 oAsnThrIleAspHisPhePheCysGluIleProProLeuLeuAlaLeu 185
 1691 CAAATCAGATTAATTAATCTTCTGTGACATCCCTTCTGATCTGTGT 1740
 185 eRysSerProValArgIleAsnGluValMetValTyrValAlaAsp1e 201
 1741 CTGTGTGAAACACTTGTCTCAATGAGTGGCCTGATTCACCTGGGCT 1790
 202 ThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyrGlyPhe11 218
 1791 TTCTATGTGTGGACTCTTCTTCTTGTATGCTACTTCTTACATTTGCAT 1840
 218 eIleValAlaIleLeuArgIleArgThrValGlnGlyLysArgLysAlaP 235
 1841 AATCTCACCACCATCTTGAGATCCAGTCCAGAGGAGGAAAGAAAGCCT 1890
 235 hSerThrCysSerSerHisLeuThrValValThrLeuTyrTyrSerPro 251
 1891 TTCTCATATGTGCTCCACCTGCTCCATGCTCTTCTTATGCGAGC 1940
 252 ValIleTyrThrTyrIleArgProAlaSerSerTyrThrPheGluArgAs 268
 1941 GCCATCTTACATATGATGACGCGCCCATCTCACTTACTCATTAAGAAAGA 1990
 268 pLysValValAlaAlaLeuTyrThrLeuValThrProThrLeuAsnProm 285
 1991 TAGCTGTGTTTCAAGTGTGTGACAGTGTGTGTACCCCATGCTAAACCTTA 2040
 285 eValTyrSerPheGlnAsnArgGluMetGlnAlaGlyIleArgLysVal 301
 2041 TAAATTACATTCATGAGGAATTAAGACATCAAGAGCTGTCAAAACATATA 2090
 seq_name: /SIDSL/gcgsdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH31995
 seq_documentation_block:
 ID AAH31995 standard; DNA; 930 BP.
 AC AAH31995;
 XX
 DT 30-JUL-2001 (first entry)
 XX
 DE Human olfactory receptor polynucleotide, SEQ ID NO: 568.
 KW Human; olfactory receptor; OR; primary scent determination;
 KW secondary scent determination; polypeptide library; odour receptor;
 KW scent profile; scent fingerprint; scent representation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200127158-A2.
 XX
 PD 19-APR-2001.
 XX
 PF 06-OCT-2000; 2000WO-US27582.
 XX
 PR 08-OCT-1999; 99US-0158615.
 XX
 PR 24-FEB-2000; 2000US-0184809.
 XX
 PA (DIGI-) DIGISCENTS.
 XX
 PA (YEDA) YEDA RES & DEV CO LTD.
 XX
 PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
 XX
 DR WPI: 2001-290713/30.
 XX
 PT New polynucleotides which encode polypeptides involved in olfactory
 XX sensation for identifying olfactory agonists and antagonists -
 XX
 PS Claim 8; Page 407; 1857pp; English.

Sequence 930 BP; 204 A; 259 C; 178 G; 289 T; 0 other;

```
alignment_block:
nc=09-975-308-9 x AAH31995
```

Align seg 1/1 to: AAH31995 from: 1 to: 930

[illegible]

```

185 erYSSerProValArgIleasnGluValMeVAlTyValAlaaspIle 201
||||| :: |||||:|||||:|||||:|||||:
563 CTTGCTCAGACACCTACATCAGTGTGGATCTGCTCTTACGTCTGTGGC 612
202 ThrIeuAlaIleGlyAspPheIleLeuThrCysIleSerTyGlyPheI 218
|||||:|||||:|||||:|||||:|||||
613 TTCATGTGAATTCAGACCACTCTCTATCAATCTTCATCTCTATACCTTAT 662
218 eIleValAlaIleLeuArgIleArgThrValGluGlyAsnGlyAlaP 235
|||||:|||||:|||||:|||||:|||||
663 CTTTGTGCAATCAGATCGCTTACCTAAAGCCCTTAAAGCTT 712
235 heSerThrCysSerSerHisLeuThrValValThrLeuTyTySerPro 251
|||||:|||||:|||||:|||||:|||||
713 TCTCACCGCGGGTCTCACTTACGGGATCATCACCCTTCATATGCGACA 762
252 ValIleTyThrTyIleArgProAlaSerSerTyThrPheGluArgAs 268
|||||:|||||:|||||:|||||:|||||
763 GTCATGTTTATGTACCTGAGCCACATCCAGCTACTCCTGGACCAAGA 812
268 pIySValValAlaIleLeuTyThrLeuValThrProThrLeuAsnProM 285
|||||:|||||:|||||:|||||:|||||
813 CAAGTGGCCTCTGTCTTACACGGTTATCAATCCCATGTTAAATCCCT 862
285 eValTySerPheGlnAsnArgIleuMetGlnaGlyIleArgIleVal 301
|||||:|||||:|||||:|||||:|||||
863 TATATCTACAGTTTGCAGAACAGATGTGAACCTGCTTCAAAAAGCTA 912

```

seq_documentation_block:

seq_documentation.docx: 957 BP
ID AAH32092 standard; DNA; 957 BP

AA AH32092
AC

AA 30-JUL-2001 (first entry)
DT

XX	Human olfactory receptor polynucleotide, SEQ ID NO: 665.
DE	

XX
XX
Human: olfactory receptor: OR: primary scent determination;

secondary scent determination; polypeptide library; odour

XX
XX

XX OS Homo sapiens.

PN WO200127158-
YY

PD 19-APR-2001.

06-OCT-2000; 2000WO-US27582.

AA 08-OCT-1999; 99US-0158615
PR

PR 24-FEB-2000; 200005-016400
XX

PA (DIGI-) DIGISCENTS.
(VEDA) VEDA RES & DEV CO LTD.

XX

DETENTION OF

XX

DR WPL; 2001-230/13/30.

PT New polynucleotides which encode polypeptides involved in sensation for identifying agonists and antagonists

XX
XX
D-23 449: 105700: English

XX	Number of isolated polynucleotides
123	1
124	1
125	1
126	1
127	1
128	1
129	1
130	1
131	1
132	1
133	1
134	1
135	1
136	1
137	1
138	1
139	1
140	1
141	1
142	1
143	1
144	1
145	1
146	1
147	1
148	1
149	1
150	1
151	1
152	1
153	1
154	1
155	1
156	1
157	1
158	1
159	1
160	1
161	1
162	1
163	1
164	1
165	1
166	1
167	1
168	1
169	1
170	1
171	1
172	1
173	1
174	1
175	1
176	1
177	1
178	1
179	1
180	1
181	1
182	1
183	1
184	1
185	1
186	1
187	1
188	1
189	1
190	1
191	1
192	1
193	1
194	1
195	1
196	1
197	1
198	1
199	1
200	1
201	1
202	1
203	1
204	1
205	1
206	1
207	1
208	1
209	1
210	1
211	1
212	1
213	1
214	1
215	1
216	1
217	1
218	1
219	1
220	1
221	1
222	1
223	1
224	1
225	1
226	1
227	1
228	1
229	1
230	1
231	1
232	1
233	1
234	1
235	1
236	1
237	1
238	1
239	1
240	1
241	1
242	1
243	1
244	1
245	1
246	1
247	1
248	1
249	1
250	1
251	1
252	1
253	1
254	1
255	1
256	1
257	1
258	1
259	1
260	1
261	1
262	1
263	1
264	1
265	1
266	1
267	1
268	1
269	1
270	1
271	1
272	1
273	1
274	1
275	1
276	1
277	1
278	1
279	1
280	1
281	1
282	1
283	1
284	1
285	1
286	1
287	1
288	1
289	1
290	1
291	1
292	1
293	1
294	1
295	1
296	1
297	1
298	1

which encode polypeptides involved in olfactory sensation. The

CC polynucleotides can be used in screening for substances, e.g.,
CC antagonists. The methods allow for the determination of primary

scent and the identification of the odour receptors used to detect it. The methods also enable determination of

secondary scents and the identification of combinations of odours

CC receptors that are involved in detecting such secondary scents.
 CC This enables the construction of a scent representation (also called
 CC a scent fingerprint or scent profile), which may be used to re-create
 CC and edit scents. Libraries of olfactory receptors are useful for
 CC determining the interaction pattern of a composition with the receptors,
 CC and can be used for determining differences in the olfactory faculties
 CC of different individuals.
 CC
 XX

50 Sequence 957 BP; 242 A; 220 C; 178 G; 317 T; 0 other;

alignment_scores:

Quality: 723.00 Length: 300
 Ratio: 3.064 Gaps: 0
 Percent Similarity: 78.667 Percent Identity: 44.667

alignment_block:

US-09-975-308-9 x AAH32092 ..

Align seg 1/1 to: AAH32092 from: 1 to: 957

```

2 AsnHisSerValValThrGluPheIleIleLeuGlyLeuThrIleValAsp18
   |||.....|.....|.....|.....|.....|.....|.....|.....|
13 AATCAACACGCTAATCAATCAATCAATCAATCAATCAATCAATCAATCAAA62
   |||.....|.....|.....|.....|.....|.....|.....|.....|
18 ogIleuGlnGlyIleIlePheLeuPhePheLeuIleValIleValIleVal35
   |||.....|.....|.....|.....|.....|.....|.....|.....|
63 TGAATTCGACATTTTACTATTCACACATCTCTTCTGACTTATTTATTCGTA112
   |||.....|.....|.....|.....|.....|.....|.....|.....|
35 IapheLeuGlyAsnMetLeuIleIleIleAlaIleValIleIleValSerAsnThr51
   |||.....|.....|.....|.....|.....|.....|.....|.....|
113 CTTGGGAGGAAATATATATATATATATATATATATATATATATATATATAT162
   |||.....|.....|.....|.....|.....|.....|.....|.....|
52 LeuHisThrPrometIleValPheLeuLeuThrLeuAlaValAlaAsp1168
   |||.....|.....|.....|.....|.....|.....|.....|.....|
163 CTGCATACACCTATGTATATTTCTAGAGCACTGGCTTTATTTATTTGACAT212
   |||.....|.....|.....|.....|.....|.....|.....|.....|
68 eIleCysThrThrSerIleIleProIlePheLeuGlyThrMetLeuThr85
   |||.....|.....|.....|.....|.....|.....|.....|.....|
213 CTGCTACACCAACCAATGATGATGATGATGATGATGATGATGATGATGATG262
   |||.....|.....|.....|.....|.....|.....|.....|.....|
85 eGluAsnThrIleSerIleValGlyCysMetSerGlnLeuPheLeuPhe101
   |||.....|.....|.....|.....|.....|.....|.....|.....|
263 AGAAAAAGCAATTTCTATGCGGGGTGTGTGTTCAACTTTTTCATTTT312
   |||.....|.....|.....|.....|.....|.....|.....|.....|
102 ThrTrpSerLeuGlyAlaGluMetValLeuPheThrThrMetAlaIle118
   |||.....|.....|.....|.....|.....|.....|.....|.....|
313 GTTTCTTTGTAGATCAGAGTGTCTCTTCTAGGCAATGCAATGCAATATGA362
   |||.....|.....|.....|.....|.....|.....|.....|.....|
118 PATGTIValAlaIleCysPheProLeuHisIleIleValIleIleMetAsn135
   |||.....|.....|.....|.....|.....|.....|.....|.....|
363 TCGTTTCATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT412
   |||.....|.....|.....|.....|.....|.....|.....|.....|
135 IHisMetCysValAlaLeuLeuSerMetValMetAlaIleAlaValThr151
   |||.....|.....|.....|.....|.....|.....|.....|.....|
413 AGGTCTATGATCAATATAGCAGCCATGCGGTGCTGCTGCTGCTGCT462
   |||.....|.....|.....|.....|.....|.....|.....|.....|
152 AsnSerTrpValHisThrAlaLeuIleMetArgLeuThrPheCysGlyPr168
   |||.....|.....|.....|.....|.....|.....|.....|.....|
463 AACTCAGTGTGATCAGTGTGATCAGTGTGATCAGTGTGATCAGTGTGAT512
   |||.....|.....|.....|.....|.....|.....|.....|.....|
168 AsnThrIleAspHisPhePheCysGluIleProIleLeuAlaIleLeu185
   |||.....|.....|.....|.....|.....|.....|.....|.....|
513 CAATCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT562
   |||.....|.....|.....|.....|.....|.....|.....|.....|
185 eCysSerProValArgIleAsnGluValMetValIleValAlaAspIle201
   |||.....|.....|.....|.....|.....|.....|.....|.....|
563 CTGTGGAAGCACTCTGCAATAGATGATGATGATGATGATGATGATGATG612
   |||.....|.....|.....|.....|.....|.....|.....|.....|
202 ThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerIleValPheIle218
   |||.....|.....|.....|.....|.....|.....|.....|.....|
613 TTCATTGTGTGACCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT662
   |||.....|.....|.....|.....|.....|.....|.....|.....|

```

```

218 eIleValAlaIleLeuArgIleArgThrValGluGlyAsnArgIleValAsp235
   |||.....|.....|.....|.....|.....|.....|.....|.....|
663 AATCTCCACCATCTTGAGATCCAGTCTCAGAGGAGAGAGAGAGAGAGCT712
   |||.....|.....|.....|.....|.....|.....|.....|.....|
235 heSerThrCysSerSerHisLeuThrValIleThrLeuThrIleSerPro251
   |||.....|.....|.....|.....|.....|.....|.....|.....|
713 TTTCTACATGTGCTCCACCTGGCCATGTCTTCTTCTTATATATATATAT762
   |||.....|.....|.....|.....|.....|.....|.....|.....|
252 ValIleThrThrIleArgProAlaSerSerIleThrPheGluValArg268
   |||.....|.....|.....|.....|.....|.....|.....|.....|
763 GCCATCTTTCATATGTACGGCCCATCTCACTTCTTATATATATATAT812
   |||.....|.....|.....|.....|.....|.....|.....|.....|
813 TAGGTGTTTCAGTGTGTACAGTGTGTATGATGATGATGATGATGATGAT862
   |||.....|.....|.....|.....|.....|.....|.....|.....|
285 eValIleSerPheGlnAsnArgIleMetGlnAlaGlyIleArgIleVal301
   |||.....|.....|.....|.....|.....|.....|.....|.....|
863 TAATTTACACATGAGATAGAGACATCAAGAAAGCTGTCAAAACTATA912
   |||.....|.....|.....|.....|.....|.....|.....|.....|

```

seqname: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32067

seq_documentation_block:

ID AAH32067 standard; DNA; 927 BP.

XX AAH32067;

XX 30-JUL-2001 (first entry)

XX Human olfactory receptor polynucleotide, SEQ ID NO: 640.

XX Human; olfactory receptor; OR; primary scent determination;

XX secondary scent determination; polypeptide library; odour receptor;

XX scent profile; scent fingerprint; scent representation; ds.

XX Homo sapiens.

XX MO200127158-A2.

XX 19-APR-2001.

XX 06-OCT-2000; 2000WO-US27582.

XX 08-OCT-1999; 99US-0158615.

XX 24-FEB-2000; 2000US-0184809.

XX (DIGI-) DIGISCENTS.

XX (YEDA) YEDA RES & DEV CO LTD.

XX Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I,

XX WPI; 2001-290713/30.

XX New polynucleotides which encode polypeptides involved in olfactory

XX sensation for identifying olfactory agonists and antagonists -

XX Claim 8: Page 437; 1857pp; English.

The present sequence is one of a number of isolated polynucleotides
 CC which encode polypeptides involved in olfactory sensation. The
 CC polynucleotides can be used in screening for olfactory agonists and
 CC antagonists. The methods allow for the determination of primary
 CC scents and the identification of the odour receptors used to detect
 CC these primary scents. The methods also enable determination of
 CC secondary scents and the identification of combinations of odour
 CC receptors that are involved in detecting such secondary scents.
 CC This enables the construction of a scent representation (also called
 CC a scent fingerprint or scent profile), which may be used to re-create
 CC and edit scents. Libraries of olfactory receptors are useful for
 CC determining the interaction pattern of a composition with the receptors,
 CC and can be used for determining differences in the olfactory faculties
 CC of different individuals.
 CC

SO Sequence 927 BP; 209 A; 253 C; 176 G; 289 T; 0 other;

alignment_scores:

Quality: 707.00 Length: 303
Ratio: 2.958 Gaps: 2
Percent Similarity: 78.878 Percent Identity: 42.244

alignment_block:
US-09-975-308-9 x AAH32067 ..

Align seg 1/1 to: AAH32067 from: 1 to: 927

```

2 AsnHisSerValValThrGluPheLeuLeuLeuGlyLeuThrLysLysPr 18
13 AATGAGAGTCCCTAATGATTCATCCCTTCTAGGCTTCACAGACCAACC 62
18 ocgLeuGlnGlyLeuLeuPheLeuPheLeuLeuValLysLeuValA 35
63 TCGTCGAGAGGCTGTCTCTTGTATTTGCTTTCTCTTCTCTCTCTGA 112
35 LArPheLeuGlnMetLeuLeuLeuLeuLeuLeuLysLysLysSerAsnThr 51
113 CCTCTGTGGGAACTTCACCATATCATCATCATCATCATCATCATCAT 162
52 LeuHisThrProMetThrValPheLeuLeuThrLeuValValAspL 68
163 CTTCATACCCCAATGATCTTTTCTCAGCAACCTCTCTTACGTGACAT 212
68 eileCysThrThrSerLleLeuProLysMetLeuGlyThrMetLeuThrs 85
213 CTGCTTACACTAGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 262
85 ercLysAsnThrLleSerThrAlaGlyCysMetSerGlnLeuPheLeuPhe 101
263 CAAGAAGAGACGATCAGTACGCTGCTGTGTGTGTGTGTGTGTGTGT 309
102 ThrTrpSerLeuGlyAla...GluMetValLeuPheThrMetAlaTy 117
310 TCTCTGCACTGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 359
117 rAspArgThrValAlaLleCysPheProLeuHisThrSerThrLleMetA 134
360 GGATCGGTACATGCTGTCTGCAAAACCCCTCACTATGATGATCATGTA 409
134 smHisMetCysValAlaLeuLeuSerMetValMetAlaLleAlaVal 150
410 ACCCAGCGCTTGTGCCAACAGCTGCACTCATCTCTCTCTCTCTCTCT 459
151 ThrAsnSerTrpValHisThrAlaLeuLleMetArgLeuThrPheCysG 167
460 GCTAGTTCCTTATCCATTCATGCACTTTTACCTTGCAATTTGCTCTGTG 509
167 yProAsnThrLleAspHisPhePheCysGluLleProLeuLeuAlaL 184
510 CAACCATAGCTGAGCATTATTTTTCGGAATGACAGCTCTCTCTCAAGT 559
184 eueCysSerProValArgLleAsnGluValMetValLysValAlaAsp 200
560 TGGCTTGTGTGGACACCATGTCATGATGATGATGATGATGATGATG 609
201 LLeuThrLeuAlaLleGlyAspPheLleLeuThrCysLleSerThrGly 217
610 GTTCTGTCTTGTGTGATTCACACGACGACGACGACGACGACGACGAC 659
217 eileLleValAlaLleLeuArgLleArgThrValGluGlyLysArgLysA 234
660 CATAAACAAGCTGTCTGTGAGGATCAATCAGTAGAGGCAAGGCAAAAG 709
234 LArPheSerThrCysSerHisLeuThrValValThrLeuThrLysSer 250
710 CCTTCAGACCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAT 759

```

```

251 ProValLleThrThrThrLleArgProAlaSerSerThrThrPheGluAr 267
760 ACCATATCTACTACTGATCGACCAACCTAGTACGATATGCCAGGACCA 809
267 gasPlysValAlaLleLeuThrLleuValThrProThrLeuAsp 284
810 AGGGAAGTTATCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 859
284 rArPheValLysSerPheGlnAsnArgGluMetGlnAlaGlyLleArgLys 300
860 CTATCATCTATCTTTAAGGACACAGATATGAAGAAGGCTCTGAGGAA 909
301 ValPheAla 303
910 CTCTCTCTCG 918
seq_name: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT.AA42407
seq_documentation_block:
ID AA42407 standard; cDNA; 1053 BP.
XX
AC AA42407;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human cDNA encoding olfactory receptor AOLF213.
XX
KW Human; olfactory receptor; G protein-coupled receptor; GPCR; odourant;
ss; food additive; cosmetic; fragrance; pharmaceutical additive.
XX
OS Homo sapiens.
XX
PN WO200168805-A2.
XX
PD 20-SEP-2001.
XX
PF 13-MAR-2001; 2001WO-US07771.
XX
PR 13-MAR-2000; 2000US-0188914.
PR 24-MAR-2000; 2000US-0192033.
PR 12-APR-2000; 2000US-0198474.
PR 24-APR-2000; 2000US-0199335.
PR 26-MAY-2000; 2000US-0207702.
PR 23-JUN-2000; 2000US-0213849.
PR 16-AUG-2000; 2000US-0226534.
PR 07-SEP-2000; 2000US-0230732.
PR 07-FEB-2001; 2001US-0266862.
XX
PA (SENO-) SENOMTX INC.
XX
PI Zozulya S;
XX
DR WPI; 2001-570867/64.
XX
P-PSDB; AAU24714.
XX
PT Nucleic acids encoding human olfactory G protein-coupled receptors,
PT useful for screening for compounds involved in olfactory sensation,
PT cosmetic industries to customise odours.
XX
PS Claim 1, Page 173; 319pp; English.
XX
XX
The invention relates to nucleic acids encoding human olfactory
receptors, OR, (a G protein-coupled receptor, GPCR). The OR's
specifically recognise molecules, odourants, that elicit specific
olfactory sensation. The human olfactory receptors and polynucleotides
encoding them are useful for screening a library of chemical compounds
for compounds that are involved in olfactory sensation. Modulators of
their activity are useful for pharmacological and genetic modulation of
olfactory signalling pathways. Therefore, they can be used in the food,
pharmaceutical and cosmetic industries to customise odours and
fragrances. The present sequence encodes a human olfactory receptor of
the invention.

```

XX
SQ Sequence 1053 BP; 234 A; 278 C; 205 G; 336 T; 0 other;

alignment_scores:
Quality: 707.00 Length: 303
Ratio: 2.958 Gaps: 2
Percent Similarity: 78.878 Percent Identity: 42.244

alignment_block:
US-09-975-308-9 x AAs42407 ..

Align seg 1/1 to: AAs42407 from: 1 to: 1053

```

2 AsnHisSerValValThrGluPheIleIleuGlyLeuThrLysLysP 18
   |||||  |||  |||  |||||  |||||  |||||  |||||  |||||
136 AATGAGAGTTCCTAATGATTCATCTCTAGGCTTCAGACCAACC 185
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
18  ogIleuGlnGlyIleIlePheLeuPhePheLeuIleValTyrLeuVal 35
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
186 TCCTCTGGAGGCTGTTCTCTTGTATTTGCTCTTCTTCTTCTTCTGA 235
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
35  labPheLeuGlyAsnMetLeuIleIleIleAlaLysIleTyrSerAsnThr 51
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
236 CCTGTGGGAACTTCACCATATCATCATCATCATCATCATCATCAT 285
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
52  LeuHisThrProMetTyrValPheLeuLeuThrLeuAlaValAlaP 68
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
286 CTTCATACCAACCAATGACTTTTCTTCAGCAACCTCTTCTTCTTCT 335
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
68  IleCysThrThrSerIleIleProLysMetLeuGlyThrMetLeuThr 85
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
336 CTGCTTCACACTAGCTCTCTCTCTCAGACCTTATGTTAAGTTGCAAG 385
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
85  erGluAsnThrIleSerTyrAlaGlyCysMetSerGlnLeuPheLeuPhe 101
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
386 CAAGAAGAGATCACTTACGCTGCTGTGTGTGTGTGTGTGTGTGTGT 432
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
102 ThrTrpSerLeuGlyAla...GluMetValLeuPheThrThrMetAla 117
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
433 TCTCTGGCACTGGGCTCCTCAGTATGATCTCTGCTGCTACACGCTT 482
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
117 rAspArgTyrValAlaIleCysPheProLeuHisTyrSerThrIleMet 134
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
483 GGATCGGTATATGCTGTCTGCAACCCCTCCATATGATGATCATATGA 532
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
134 snHisMetCysValAlaLeuLeuSerMetValMetAlaIleAlaVal 150
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
533 ACCCAGCGCTTGTGCAACACCTGTCATCTCTGCTCAGCGCTTGTG 582
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
151 ThrAsnSerTyrValHisThrAlaIleuIleMetArgLeuThrPheCys 167
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
583 GCTAGTTCCTTAATCCATGCACTTATACCTTCAATGCTCTCTGTGG 632
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
167 yProAsnThrIleAspHisPhePheCysGluIleProLeuLeuAla 184
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
633 CAACCATAGGCTGACCATTTATTTGCGAAGTACACGCTCTCTCAAGT 682
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
184 euSerCysSerProValArgIleAsnGlnValMetValTyrValAlaAsp 200
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
683 TGCTTGTGTGGACACCATGTCATGAATGTGTGTGTGTGTGTGTGT 732
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
201 IleThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyrGly 217
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
733 GTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 782
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
217 eileIleValAlaIleLeuArgIleArgThrValIleGlyLysArgLys 234
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
783 CATAACTCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 832
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
234 labSerThrCysSerSerHisLeuThrValAlaThrLeuTyrTyrSer 250
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
833 CCTTCAGACCTGCTCTCTCCACCTTACAGTGTGTGTGTGTGTGTGT 882

```

```

251 ProValIleTyrThrTyrIleArgProAlaSerSerTyrThrPheGlu 267
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
883 ACCATATCTACGTGTACCTGCAACCTGATGACATGATGCGGAGCA 932
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
267 gAspLysValValAlaIleuTyrThrLeuValThrProThrLeuAsn 284
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
933 AGGGAAGTTATCTCCCTCTTCTACACCATGTATACCCCTTTAATC 982
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
284 roMetValTyrSerPheGlnAsnArgGluMetGlnAlaGlyIleArgLys 300
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
983 CTATCATCTATACCTTTAAGAAACAAGATGAAAGAAGCGCTGTGAGG 1032
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
301 ValPheAla 303
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1033 CTTCCTCTCG 1041

```

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32033

seq_documentation_block:
ID AAH32033 standard; DNA; 927 BP.

XX AAH32033;

DT 30-JUL-2001 (first entry)

DE Human olfactory receptor polynucleotide, SEQ ID NO: 606.

XX Human; olfactory receptor; OR; primary scent determination;

KW secondary scent determination; polypeptide library; odour receptor;

KW scent profile; scent fingerprint; scent representation; ds.

XX Homo sapiens.

PN WO200127158-A2.

PD 19-APR-2001.

PF 06-OCT-2000; 2000WO-US27582.

PR 08-OCT-1999; 99US-0158615.

PR 24-FEB-2000; 2000US-0184809.

PA (DIGI-) DIGISCENTS.

PA (YEDA) YEDA RES & DEV CO LTD.

PI Bellenson J, Smith D, Lancel D, Glusman G, Fuchs T, Yanai I;

DR WPI; 2001-290713/30.

PT New polynucleotides which encode polypeptides involved in olfactory

PS sensation for identifying olfactory agonists and antagonists -

XX Claim 8: Page 423; 1857pp; English.

The present sequence is one of a number of isolated polynucleotides which encode polypeptides involved in olfactory sensation. The polynucleotides can be used in screening for olfactory agonists and antagonists. The methods allow for the determination of primary scents and the identification of the odour receptors used to detect these primary scents. The methods also enable determination of secondary scents and the identification of combinations of odour receptors that are involved in detecting such secondary scents. This enables the construction of a scent representation (also called a scent fingerprint or scent profile), which may be used to re-create and edit scents. Libraries of olfactory receptors are useful for determining the interaction pattern of a composition with the receptors, and can be used for determining differences in the olfactory faculties of different individuals.

SQ Sequence 927 BP; 201 A; 278 C; 166 G; 282 T; 0 other;

alignment_scores:

Quality: 706.00 Length: 305
Ratio: 2.954 Gaps: 4
Percent Similarity: 78.361 Percent Identity: 44.590

alignment_block:

US-09-975-308-9 x AAS42313 ..

Align seg 1/1 to: AAS42313 from: 1 to: 930

```

1 MetAsnHis...SerValValThrGluPheIleIleuGlyLeuThrly 16
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
10 ATCAACACACACAGCAGTGTCTCGAGTTATCTCTCGGAGCCTTCCTC 59
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
16 slySPROGluLeuGlnGlyIleIlePheLeuPheLeuIleValylr 33
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
60 CCGGCCGTGAGGACCAAAAGACACTCTTGTCTCTCTCTCATGCTGACC 109
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
33 euValAlaPheLeuGlyAsnMetLeuIleIleIleAlaIleValylr 49
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
110 TGCTCACCATTACAGGACCTGCTCATCTCTGCGCATTCGCTTAC 159
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
50 AsnThrLeuHisThrPrometlyrValPheLeuLeuThrLeuAlaVal 66
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
160 CCCCATCTTCAGACCCCTATGTATCTCTCTGAGTTTCTGCTCTCAC 209
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
66 IAspIleIleCysThrThrSerIleIleProlysmetLeuGlyThrmet 83
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
210 TGAATTTGCTTTACACACAGCGTGTGCCCCAAGATGCTATGAACTTCC 259
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
83 eutHrSerGluAsnThrIleSerTyralGlyCysmetSerGlnLeuPhe 99
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
260 TGTCAGAAAGAGACCATCTCTATGCTGGGTGTGACACAGATGTAT 309
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
100 LeuPheThrTrpSerleuGly...AlaGluMetValLeuPheThrThrMe 115
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
310 ...TTTCTCTATGCTGTGGCAGACAGTACAGCTGCTTGTGACAGTCAT 356
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
115 tAlaTyraSPArgTyraIleAlaIleCysPheProLeuHisThrSerhri 132
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
357 GGGCTTTGACCGGTATGTGGCGTGTGACCTTCCACTATGTGACCA 406
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
132 leMetAsnHisIleMetCysValAlaLeuLeuSerMetValMetAlaIle 148
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
407 CCATGAGCCACACACATGTGTCTGTGTGTGTGTCTCTCTCTCTCTT 456
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
149 AlaValThrAsnSerTrpValHisThrAlaLeuIleMetArgLeuThrph 165
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
457 CCTCACCCTCCACTCCTCGACACACTTGTGTGATGCTGCTCACCTT 506
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
165 eCysGlyProAsnThrIleAspHisPheCysGluIleProLeuLeu 182
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
507 CTGTGACTCCAAATGTTATCCACACTTCTGTGACCTCAGCCCTGTGC 556
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
182 euAlaLeuSerCysSerProValArgIleAsnGluValMetValTyraVal 198
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
557 TGAATTTGCTCTCTCTCTCATATTGTCTCAATGAAATGTGCGAGATGACA 606
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
199 AlaAspIleThrLeuAlaIleGlyAspPheIleLeuThrIleSerly 215
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
607 GAAGCACCCTATGTTGTGTGATGCTGTTTCTGTGATGCTTCTCTCTTA 656
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
215 rGlyPheIleIleValAlaIleLeuArgIleArgThrValGluGlyAsa 232
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
657 TATACAGATTCCTACAGATTCCTCAAGATTCCTACTTCTGCGGAAC 706
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
232 rGlyAlaPheSerThrCysSerSerHisLeuThrValValThrLeuTy 248
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
707 GCAAGACCTTCTCACCCTGTGTGTTTACTCTCAACGCTGTGAGCGCTTT 756
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
249 TyrSerProValIleTyThrTyThrIleArgProAlaSerSerTyThrph 265
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||

```

```

757 TATGAGACATCTCTGCTGCTATTATTAAGCCCCCATTCACACTACGCTGT 806
265 eGluArgAspLysValValAlaIleLeuTyThrLeuValThrProThrL 282
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
807 C...AAGGACCCAGCTGGCAACAAATGTTTACACAGTTTGTTCATTCATG 853
282 euAsnProMetValTyThrSerPheGlnAsnArgGluMetGlnAlaGlyLe 298
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
854 TCATCTCTTTTATCTACAGCTGTGAGAAACAAAGACTGAAACAGGCGCTG 903
299 ArgLysValPheAla 303
   ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
904 AGAAGCTTATGAGC 918

```

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NR2001A.DAT:AAH32140

```

seq_documentation_block:
ID AAH32140 standard; DNA; 948 BP.
XX
AC AAH32140;
XX
DT 30-JUL-2001 (first entry)
XX
DE Human olfactory receptor polynucleotide, SEQ ID NO: 713.
XX
KW Human: olfactory receptor; OR; primary scent determination;
XX secondary scent determination; polypeptide library; odour receptor;
XX scent profile; scent fingerprint; scent representation; ds.
XX
OS Homo sapiens.
XX
PN WO200127158-A2.
XX
PD 19-APR-2001.
XX
PF 06-OCT-2000; 2000WO-US27582.
XX
PR 08-OCT-1999; 99US-0158615.
XX
PR 24-FEB-2000; 2000US-0184809.
XX
PA (DIGI-) DIGISCENTS.
XX
PA (YEDA ) YEDA RES & DEV CO LTD.
XX
PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
XX WPI; 2001-290713/30.
XX
PT New polynucleotides which encode polypeptides involved in olfactory
XX sensation for identifying olfactory agonists and antagonists -
XX
PS Claim 8: Page 467, 1857pp; English.
XX
CC The present sequence is one of a number of isolated polynucleotides
XX which encode polypeptides involved in olfactory sensation. The
XX polynucleotides can be used in screening for olfactory agonists and
XX antagonists. The methods allow for the determination of primary
XX scents and the identification of the odour receptors used to detect
XX these primary scents. The methods also enable determination of
XX secondary scents and the identification of combinations of odour
XX receptors that are involved in detecting such secondary scents.
XX CC This enables the construction of a scent representation (also called
XX a scent fingerprint or scent profile), which may be used to re-create
XX and edit scents. Libraries of olfactory receptors are useful for
XX determining the interaction pattern of a composition with the receptors,
XX and can be used for determining differences in the olfactory faculties
XX of different individuals.
XX
SQ Sequence 948 BP; 240 A; 217 C; 162 G; 329 T; 0 other;

```

alignment_scores: Quality: 699.00 Length: 300
Ratio: 2.937 Gaps: 0

Percent Similarity: 79.333 Percent Identity: 42.333

alignment_block:

US-09-975-308-9 x AAH32140 ..

Align seg 1/1 to: AAH32140 from: 1 to: 948

```
2 AsnHisSerValValThrGluPheIleIleLeuGlyLeuThrLysLysPr 18
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13 AATTTACACCACTGCCACTGAGTCTTATTTGTTGATTCACAGATTATCT 62
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18 oGluLeuGlnGlyIleIlePheLeuPhePheLeuIleValIleLeuVal 35
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
63 ACCCTCTAGAGTCCACACTGTTCTTGATTCCTTCTGTATATACATTA 112
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
35 IapheLeuGlyAsnMetLeuIleIleIleAlaLysIleIleSerAsnThr 51
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
113 CTATGCTCGGAAATATCTTAAATCTTAAATCTTAAATATTAATCAAGC 162
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
52 LeuHisThrProMetTyrValPheLeuLeuThrLeuAlaValAlaSpr1 68
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
163 CTTCAATTCCTCATGCTATATTTCTTATAGCAACTATCTTCTTAGACAT 212
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
68 eIleCysThrThrSerIleIleProLysMetLeuGlyThrMetLeuThr 85
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
213 CAGCTGTTCTACAGCAATCACTCTAAATGCTGGCAAACTCTTGGCAT 262
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
85 eArgLysAsnThrIleSerTyrAlaGlyCysMetSerGlnLeuPheLeuPhe 101
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
263 CCAGGAAAGCATCTCTCTTATGGGTGTCACATCAAAATGTTTCTCTTC 312
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
102 ThrThrSerLeuGlyAlaGluMetValLeuPheThrThrMetAlaTyrAs 118
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
313 GCTTCTTTGCTGATGCTGAGTGCCTTATCTCGCAGCAATGCTTATGA 362
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
118 PargTyrValAlaIleCysPheProLeuHisTyrSerThrIleMetAsnH 135
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
363 CCGCTATGAGCATCTGCTCATGACCCACTGCTGATCTACACTGATGCTCA 412
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
135 LshIleMetCysValAlaLeuLeuSerMetValMetAlaIleAlaValThr 151
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
413 GGAGAGTCTGCTGCTTCATTTGTTGGCATATTTTCAGTGGAAAGTACA 462
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
152 AsnSerThrPValHisThrAlaLeuIleMetArgLeuThrPheCysGlyPr 168
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
463 ACATCTCTGCTCATGTCCTCATCATTCAGCTGTCATTTTGGGCTC 512
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
168 oAsnThrIleAspHisPhePheCysGluIleProProLeuLeuAlaLeus 185
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
513 CAATATCGTCAATCATTTTCTGTGATATCCACCTCTTGGCTTTAT 562
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
185 eCysSerProValArgIleAsnGluValMetValTyrValAlaAspIle 201
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
563 CATGTACAGACACTCAGATCAACACACTCTGCTTTCCTTGTGCAGC 612
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
613 ThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyrGlyPheI 218
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
218 eIleValAlaIleLeuArgIleArgThrValGluGlyLysArgLysAlaP 235
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
663 CCTCATCATGCTGTGAGCATCAAGTCTCAGGTGGCAGAGCAAAACAT 712
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
235 heseThrCysSerSerHisLeuThrValValThrLeuTyrTyrSerPro 251
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
713 TCTCCACTTGTGCTCCACCTCATACAGTCACTATCTATGTGAGCG 762
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
252 ValIleTyrThrTyrIleArgProAlaSerSerTyrThrPheGluArg 268
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
763 CTCTGTTTATGTACTACAGCCACCATAGTATTTCCCTAGACACTGA 812
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
268 pLysValValAlaIleLeuTyrThrLeuValThrProThrLeuAsnPro 285
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```
813 TAAGTGTGGCAGTCTTTTATCTGTGATTTTCCCATGTTTAAATCCAA 862
285 eValTyrSerPheGlnAsnArgGluMetGlnAlaGlyIleArgLysVal 301
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
863 TATTTATAGTTTCAGAAACAAGGATGTGAAATAATGCTCTCAAAAAGCTA 912
```

OM of: US-09-975-308-9 to: EST:* out_format : pfs

Date: Jun 24, 2002 12:23 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame+22n.model -DEV=xlh
-O=/gen2.1/SPOT/spool/US09975308/runat_2062002_090302_26225/app_query.fasta_1.368
-DB=EST -QFMT=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000
-MINMAP=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -GAPOP=6.000
-OGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DEEXT=7.000 -START=1 -MATRIX=blomsm62 -TRANS=humad40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZ=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09975308.@CGN1_1.6099
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPHY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-975-308-9
Query length: 307
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 1680.040000
```

score_list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
gb.htc:BC016940	+	633.00	1669.28	2.6e-50	2021	BC016940 Homo sapiens, Stimulat
gb.gss:AZ269227	+	553.00	942.42	3.0e-43	642	AZ269227 2M0241274R Mouse 10kb
gb.gss:BH331857	+	543.50	926.78	2.2e-42	853	BH331857 CH230-46N15.TV CHOR1-2
gb.htc:AK016560	+	543.00	910.26	1.9e-41	3063	AK016560 Mus musculus adult ma
gb.htc:AK017036	+	542.00	916.06	8.9e-42	1394	AK017036 Mus musculus adult ma
gb.gss:AO428256	+	531.00	907.49	2.7e-41	470	AO428256 CTRP1-EI-2578P1.TF CT
gb.gss:AZ909618	+	516.50	877.47	1.3e-39	797	AZ909618 RPTC-24-222618.TJ RPTC
gb.gss:AZ913339	+	514.00	873.59	2.1e-39	762	AZ913339 RST12467 Athysys RAE
gb.gss:AZ913406	+	511.50	868.89	3.8e-39	794	AZ913406 RPTC-24-165C12.TV RPTC
gb.gss:AZ593814	+	504.00	858.97	1.3e-38	580	AZ593814 IM0405K03R Mouse 10kb
gb.gss:AZ255734	+	504.00	856.82	1.8e-38	726	AZ255734 RPTC-23-16619.TJ RPTC
gb.est2:BG197640	+	503.50	855.08	2.2e-38	796	BG197640 RST17016 Athysys RAE
gb.gss:AZ607393	+	501.00	851.47	3.5e-38	740	AZ607393 IM0429N15R Mouse 10kb
gb.gss:AZ53949	+	493.50	839.42	1.6e-37	676	AZ53949 IM0093J01F Mouse 10kb
gb.gss:BH111204	+	483.00	819.02	1.7e-36	698	BH111204 RPTC-24-331A10.TVB RPTC
gb.gss:AZ380178	+	482.50	818.11	2.5e-36	660	AZ380178 IM0136A04F Mouse 10kb
gb.gss:AZ086625	+	478.50	813.43	4.6e-36	686	AZ086625 RPTC-23-26F4.TV RPTC-2
gb.gss:AZ709687	+	473.50	804.90	1.4e-35	680	AZ709687 RPTC-24-82E2.TJ RPTC-2
gb.gss:AZ574496	+	469.00	796.94	3.8e-35	695	AZ574496 2M0249B07F Mouse 10kb
gb.gss:AZ103967	+	466.00	793.01	6.3e-35	611	AZ103967 RPTC-23-33618.TV RPTC-1
gb.htc:AK016338	+	464.00	780.97	3.0e-34	1501	AK016338 Mus musculus adult ma
gb.gss:BH272774	+	456.00	773.14	8.1e-34	805	BH272774 CH230-56L1.TJ CHOR1-23
gb.gss:AZ638594	+	453.00	770.08	1.2e-33	646	AZ638594 IM0498M06R Mouse 10kb
gb.gss:AZ911698	+	452.50	769.00	1.4e-33	661	AZ911698 RPTC-23-166A20.TV RPTC
gb.gss:AZ555190	+	452.00	767.60	1.6e-33	699	AZ555190 RPTC-23-179A13.TV RPTC
gb.gss:BH078380	+	451.00	766.48	1.9e-33	656	BH078380 RPTC-24-34E9.TV RPTC-23
gb.gss:AZ111477	+	447.50	766.11	2.0e-33	682	AZ111477 RPTC-23-9A5.TJ RPTC-23
gb.gss:BM337934	+	442.00	751.30	1.3e-32	634	BM337934 CH230-111E1.TV CHOR1-2
gb.gss:AZ086388	+	439.00	744.62	3.2e-32	743	AZ086388 CH230-1976F.TJ CHOR1-2
gb.gss:AZ090606	+	436.50	740.09	5.0e-32	692	AZ090606 RPTC-22-27B20.TV RPTC-1
gb.gss:BM342053	+	432.50	736.07	9.4e-32	616	BM342053 CH230-51L13.TV CHOR1-2
gb.gss:AZ725747	+	432.50	732.23	1.5e-31	642	AZ725747 RPTC-23-9B112.TVB RPTC-2
gb.gss:AZ235331	+	431.50	732.55	1.5e-31	680	AZ235331 RPTC-23-69A15.TVB RPTC
gb.est2:BG068751	+	431.00	730.98	1.8e-31	732	BG068751 H3068H10-3 N1A Mouse 1

```
gb.gss:BM078443 - 430.00 730.56 1.9e-31 639 | BH078443 RPTC-24-349G9.TV RAE
gb.gss:BM359995 + 428.50 726.37 3.3e-31 756 | BH359995 CH230-12616.TJ CHOR
gb.gss:AZ392036 + 428.00 727.32 2.9e-31 625 | AZ392036 IM0154P24F Mouse 10
gb.gss:AZ972672 - 427.00 725.20 3.8e-31 652 | AZ972672 2M0246D10R Mouse 10
gb.gss:BH291047 - 426.50 723.51 4.7e-31 711 | BH291047 CH230-103B15.TJ CHC
```

seq_name: gb.htc:BC016940

seq_documentation_block:

```
LOCUS BC016940 2021 bp mRNA linear HTC 09-NOV-2001
DEFINITION Homo sapiens, similar to olfactory receptor, family 2, subfamily A, member 4, clone IMAGE:4424116, mRNA.
ACCESSION BC016940
VERSION BC016940.1 GI:16877381
KEYWORDS HTC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2021)
```

REFERENCE

```
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
```

REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

COMMENT

```
Contact: MGC help desk
Email: cgaps-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing School of Medicine, Stanford Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mdc@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.
```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>

Series: IRM Plate: 27 Row: 1 Column: 22
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency OR analysis, Genomscan gene prediction, Similarity but not identity to protein

This clone has the following problem: frame shifted.

Location/Qualifiers

FEATURES

```
1..2021
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4424116"
/tissue_type="Kidney, hypernephroma"
/clone_lib="NHL_MGC_89"
/lab_host="DH10B"
/note="Vector: pCMV-sport6"
```

BASE COUNT

470 a 546 c 427 g 578 t

ORIGIN

```
alignment_scores: 633.00 Length: 314
Ratio: 2.839 Gaps: 3
Percent Similarity: 71.019 Percent Identity: 41.401
```

alignment_block:

US-09-975-308-9 x BC016940

Align seg 1/1 to: BC016940 from: 1 to: 2021

```
2 AsnHsserValValThrGluPheIleIleLeuGlyLeuThrLysLysPr 18
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
438 AATGAGCAATGTCACAGACTTCCTCCTACTGGGATTTCCCGGGCCC 487
```

```

18  ogluLeuGlnGlyIleIlePheLeuPhePheLeuIleValIleTyrLeuVala 35
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
488  AAGATGATGATGCTCTTGGCTCTTCCCTGTTATGATCTTCA 537
35  laPheLeuGlyAsnMetLeuIleIleIleAlaLysIleTyrSerAsnThr 51
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
538  CCGTGGTGGGAATGGACCAATCCCTGGGCTCACTCAGACTCCAGA 587
52  LeuIleThrProMetTyrValPheLeuLeuThrLeuAlaValaIlePhe 68
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
588  CTCACACACCCCATGTAATCTTCTCTCACAACCTGGCCGTCACACAT 637
68  eileCysThrThrSerIleIlePhePheLeuGlyIleThrMetLeuThr 85
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
638  CGCTATGCTGTCACACACATGCTGCTGCTGCTGCTGCTGCTGCTG 687
85  egluAsnThrIleSerTyrAlaGlyCysMetSerGlnLeuPheLeuPhe 101
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
688  CAGCCACACCCCATGTAATCTTCTCTCACAACCTGGCCGTCACACAT 737
102  ThrTrpSerLeuGlyAlaGlyMetValLeuPheThrThrMetAlaTyrAs 118
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
738  TTGAGTTTGCACATGTAATGCTCTGTTGGTGGTGGTGGTGGTGGT 787
118  PArgTyrValAlaIleCysPheProLeuIleIleTyrSerThrIleMetAsn 135
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
788  TCGGACCTGGCCATCTGCACTCTCCGATTTTCAATCAATCAATGACCT 837
135  LSHMetCysValAlaLeuLeuSerMetValMetAlaIleAlaValThr 151
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
838  GGAAGTGTGATCATCTG.....GCCATGACT 866
152  AsnSerTrp.....ValHisThrAlaLeuIle 160
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
867  ..TCCTGGACATGAGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 913
160  eMeArgLeuThrPheCysGlyProAsnThrIleAspHisPhePheCysG 177
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
914  CCTAAGATGCTCTTGGGCTCTGGAATCAACCATCTCTCTG 963
177  LuIleProLeuLeuAlaLeuSerCysSerProValArgIleAsnGlu 193
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
964  AATCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1013
194  ValMetValTyrValAlaAspIleThrLeuAlaIleGlyAspPheIle 210
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1014  GGGGTCATCTTGGACCTGGCATGTTCACTGGTGGGACCTGCTGCT 1063
210  uThrCysIleSerTyrGlyPheIleIleValAlaIleLeuArgIleArg 227
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1064  GGTGCTGGTCTCTACTACATCTGCGGCGCATCTGAGATCCAGT 1113
227  hValGluGlyLysArgLysAlaPheSerThrCysSerSerHisLeuThr 243
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1114  CGGGGAGGGCGCGAAGAGCCCTTCTCACCCTGCTCTCCACCTCTGC 1163
244  ValValIleLeuTyrTyrSerProValIleTyrThrTyrIleArgProAl 260
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1164  GTAGTGGGCTCTTCTTGGCAGCGGCATGTCATGATGATGCCCCCTAA 1213
260  aSerSerTyrThrPheGluArgAspLysValAlaAlaLeuTyrThrL 277
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1214  GTCCCGCCATCTGAGAGCAGACAGAGGCTCTTTTCTATTTTACAGTT 1263
277  euValIleThrProThrLeuAsnProMetValTyrSerPheGlnAsnArg 293
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1264  CTTTCAACCCGATGCTAAACCCCTGATTTTCAACCTGAGGAATGTAGAG 1313
294  MetGlnAlaGlyIleArgLysValPheAlaPheLeuIleHis 307
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1314  GTCAAGGTGCTCTGAGAGACACTGTGCAAGGAATGATCAT 1355

```

seq_name: gb_gss:A2969227

seq_documentation_block:

LOCUS A2969227 642 bp DNA linear GSS 27-APR-2001
 DEFINITION 2M0241J24R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
 clone UUGC2M0241J24 R, DNA sequence.

ACCESSION A2969227
 VERSION A2969227.1 GI:13840454

KEYWORDS

GSS..

SOURCE

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 642)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meene,E., Pedersen,T., Reilly,
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D., Weiss,R.

AUTHORS

Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts

TITLE

Unpublished (2000)

COMMENT

Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0241 row: J column: 24
 Seq primer: CACACAGGAAGACGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 642.
 Location/Qualifiers

FEATURES

source

1..642

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC2M0241J24"

/clone_lib="Mouse 10kb plasmid UUGC2M library"

/sex="Female"

/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PMD29uv, Purified genomic DNA from M.
 musculus C57BL/6J (female) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD2 (g114732114[bp]AP129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT
 ORIGIN

130 a 166 c 129 g 217 t

alignment_scores:

Quality: 553.00 Length: 213
 Ratio: 3.331 Gaps: 0
 Percent Similarity: 77.934 Percent Identity: 47.887

alignment_block:

US-09-975-308-9 x A2969227

Align seg 1/1 to: AZ69227 from: 1 to: 642

```

76 Prolysmetleucllythmetleuthrsergluanshrilserlyal 92
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
   3 CCCAGATGATGGTGCATCTCTGTCAGAGAAATCATTTCCATGG 52
92 aglyCysMetSerGlnPheLeuPheThrTrpSerLeuGlyAlaGlu 109
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
53 AGGCTGTGACCCAGCTCTTTCATTCATTTCTTTGCTGCTCAGAGT 102
109 etValMetPheThrThrMetAlaTyraSparGlyValAlaIleCysPhe 125
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
103 GTCCTCCTCGGACGACATGATGATGATATATGCTATCTGTAG 152
126 ProLeuHisIySerThrIleMetAsnHisMetCysValAlaLeu 142
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
153 CCGTTAAGGTACCTATTATGAAACAGGCTGTGCTGCTGTTAGC 202
142 userMetValMetAlaIleAlaValThrAsnSerTrpValHisThrAla 159
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
203 AGCTTCATGCTGGACAGTGGTTCCTCACTCACTGCTGACACGGTTT 252
159 euileMetArgLeuThrPheCysGlyProAsnThrIleAspHisPhe 175
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
253 TGACCTTCACCGCCCTTTTGTGTACACATCAGATCATATTTCTTC 302
176 CysGluIleProProLeuLeuAlaLeuSerCysSerProValArgIleAs 192
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
303 TGTGACATACCTCCCTGTCATCTGTGTGTGTGTGTGTGTGTGTGT 352
192 nguValMetValTyraValAlaAspIleThrLeuAlaIleGlyAspPhe 209
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
353 TGACATGCTGTGCTGTCATTCATTCGATCCATAGCTGATGCTCTTC 402
209 leuThrCysIleSerTyrglyPheIleIleValAlaIleLeuArgIle 225
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
403 TGTGCGTCATCCCTTCCTACCTTACATCATCTCCACATCCGAGAGATC 452
226 ArgThrValGluGlyArgIleAlaPheSerThrCysSerSerHisIle 242
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
453 CGTTCCTCGAGGGGAGGACCAAGGCTTTCCACCTGTGCTCCACCT 502
242 uThrValValThrLeuTyrrTyrrSerProValIleTyrrThrIleArgP 259
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
503 GCTCATGTTATCTCTATTATGCGAGTCTATCTTACGATGATGAGGC 552
259 roAlaSerSerTyrrPheGluArgAspIleValAlaAlaLeuTyrr 275
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
553 CCATCTCATCTTACTCTCTAGAGAAAGATGATGATCTCAGTGTCTAT 602
276 ThrLeuValThrProThrLeuAsnProMetValTyrrSer 288
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
603 AGTGTGTGCACACCCATGCTGAATCTGTAAATTTATACG 641

```

seq_name: gb_gss:BH331857

seq_documentation_block:

LOCUS BH331857 853 bp DNA linear GSS 03-DEC-2001
 DEFINITION CH230-46N15.TV CHOR1-230 Segment 1 Rattus norvegicus genomic clone
 CH230-46N15, DNA sequence.

ACCESSION BH331857
 VERSION BH331857.1 GI:17262571

KEYWORDS GSS.

SOURCE

ORGANISM Norway rat.
 Rattus norvegicus
 Mammalia; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Eumetazoa; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 853)

AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
 A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
 Jong,P. and Fraser,C.M.

TITLE Rat BAC End Sequences from Library CHOR1-230 EcORI segment
 JOURNAL Unpublished (1999)
 COMMENT Other_GSSs: CH230-46N15.TJ

Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org

Clones are derived from the rat BAC library CHOR1-230
 (<http://www.chori.org/bacpac/rat230.htm>). For BAC library
 availability, please contact Pliet de Jong (pdejong@mail.cno.org).
 Clones may be purchased from BACPAC Resources
 (http://www.chori.org/bacpac/orceting_information.htm). BAC end
 page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
 Plate: 46 row: N column: 15
 Seq primer: T7
 Class: BAC ends.

FEATURES
 source location/Qualifiers

1..853
 /organism="Rattus norvegicus"
 /strain="BN/SSHsd/MCM"
 /db_xref="taxon:10116"
 /clone="CH230-46N15"
 /clone_lib="CHOR1-230 Segment 1"
 /sex="female"
 /cell_type="Brain"
 /note="Vector: pTARBAC2.1; Site_1: EcORI; Site_2: EcORI;
 CHOR1-230 Rat (BN/SSHsd/MCM) BAC library produced by
 Pliet de Jong"
 BASE COUNT 210 a 172 c 158 g 313 t
 ORIGIN

alignment_scores: Quality: 545.50 Length: 290
 Ratio: 2.687 Gaps: 5
 Percent Similarity: 70.000 Percent Identity: 38.966

alignment_block:
 US-09-975-308-9 x BH331857 ..

Align seg 1/1 to: BH331857 from: 1 to: 853

```

13 GlyLeuThrIlyLysProGluLeuGlnGlyIleIlePheLeuPhePhe 29
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
   3 GGGACATTACGATACACGAACTCAGT..... 29
29 uIleValTyrrLeuValAlaPhe.....LeuGlyA 39
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
30 CTTGCTCATTTGAGCCCTTCTATCATATATGTATACAGTGTGGGA 78
39 smetLeuIleIleIleAlaIleAlaIleTyrrSerAnthrIleuHisThrPro 55
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
79 ACATTCATCAATACATGCGCTGCTTGTATTCACCACTCCATACACCT 128
56 MetTyrrValPheLeuLeuThrIleuAlaValAlaAspIleIleCysThrTh 72
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
129 ATGTACTCTCTTCATTCACCACTGCTTTCATGACATCTGTTTACAC 178
72 rserIleIlePheLeuMetLeuGlyThrMetLeuThrSerGluAsnThrI 89
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
179 CACCACCATCCCAAAATGTTGGTGAATATTCAGACTCAGATCAGTCCA 228
89 leSerTyrrAlaGlyCysMetSerGlnLeuPheLeuPheThrTrpSerLeu 105
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
229 TTATGTATATGAGCTCCACACCAATATCGCTTTTGTAGCTTTGCA 278
106 GlyAlaGluMetValLeuPheThrThrMetAlaTyraSparGlyValAla 122
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
279 GGATTTGAAATGGAATTCGTGAATGATGGCTCATGATAGGTTTGTGCG 328

```

122 atleCyspheProLeuHstYrSerThr11eMetasHstMetCysV 139
 139 alaleuSerMetValMetAl1leAlaValThrsanSerTrpVal 155
 379 GGTGATGCTTTGTTGCTCTTCTTGTATAGCAATTCAGTCTTGCTC 428
 156 HstThrAlaLeu1leMetArgLeuThrPheCysGlyProAsnThr1leas 172
 429 CACACTTGTATGGCAGCTGCGCTGTCTCATCTTCGCAAGAGTGGAATCC 478
 172 PhsPhePheCysGlu1leProProLeuAlaLeuSerCysSerProV 189
 479 TCACCTTTTGTGAACTAGCTCATATTCGAACTGCAAGTGTTCACATA 528
 189 alArg1leasngluValMetValTyrValAlaAsp1leThr1leAla1le 205
 529 TTCTATCATATATATTCGTGTATTTG.....GTGACAGATCTGTG 572
 206 GlyAspPhe1leLeuThrCys1leSerTyrGlyPhe.....1le1leVa 220
 573 GGTGCTTCTCCACTGCTGTGTATATTTATCTTATCTAGTAATTTATTC 622
 220 1Ala1leLeuArg1leArgThrValGluGlyValArgGlyAlaPheSerT 237
 623 CTCTGTTTAAATTCATCATCACTGACGAGAAATATATAAGTTTCTCCA 672
 237 hrcYsSerSerHst1leuThrValValThr1leuTyrSerProVal1le 253
 673 CAGTGTCTCACACTAGTGTAGTGTATCTGTATCTGTATGTACAGTTT 722
 254 TyrThrThr1leArgProAlaSerSerTyrThrPheGluArgAspLysVa 270
 723 GGTGTTATCTGAGTCTGCTGGCAGTCATAGCCTGAAAGATGCAAT 772
 270 1Val1Ala1leuThr1leuThrLeuValThrProThrLeuAsnProMetVal 286
 773 AGCCTGTGTATGTACACTGTGTGTCACTCCATGATGATGATCCCTTATA 822
 287 TysSerPheGlnAsnArg 292
 823 TATAGCTTTGAGATTAAG 840

seq_name: gb_hlc:AK016560

seq_documentation_block:

LOCUS AK016560 3063 bp mRNA linear HTC 19-JAN-2002
 DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched
 library, clone:4932441H21:similar to T1 OLFACTORY RECEPTOR, full
 insert sequence.

ACCESSION AK016560
 VERSION AK016560.1 GI:12855357
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,
 clone.lib:RIKEN full-length enriched mouse cDNA library
 clone:4932441H21.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)

PUBMED 99279253

MEDLINE 10349636

2 (sites)

REFERENCE

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20499374

PUBMED 11042159
 REFERENCE 3 (sites)
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Kono, H., Akiyama, J., Nishi, K., Katsuna, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasliwal, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multiplexed sequencer
 Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL 20030913
 MEDLINE 11076861
 4 (sites)

REFERENCE The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5 (bases 1 to 3063)

ADACHI, J., AIZAWA, K., AKAHIRA, S., AKIMURA, T., AONO, H., ARAI, A.,
 ARAKAWA, T., BALDARELLI, R., BONO, H., BROWNSTEIN, M., BULL, C.,
 CARNINCI, P., FUKUDA, S., FUKUNISHI, Y., FURUNO, M., HANAGAKI, T.,
 HARA, A., HAYATSU, N., HILL, D., HIMOTO, K., HIROKA, T., HORI, F.,
 HUNE, D., IMOTANI, K., ISHII, Y., ITOH, M., IZAWA, M., KASUKAWA, T.,
 KATO, H., KAWAI, J., KOJIMA, Y., KONNO, H., KODA, M., KOYA, S.,
 KURIHARA, C., MATSUYAMA, T., MIYAZAKI, A., NISHI, K., NOMURA, K.,
 NUMAZAKI, R., OHNO, M., OKAZAKI, Y., OKIDO, T., OWA, C., QUACKENBUSH, J.,
 SATO, H., SATO, R., SAKAI, C., SAKAI, K., SANO, H., SASAKI, D.,
 SCHIMM, L., SHIBATA, K., SHIBATA, Y., SHINGAWA, A., SHIRAKI, T.,
 SOGABE, Y., SUZUKI, H., TAGAMI, M., TAGAWA, A., TAKAHASHI, F.,
 TANAKA, T., TEJIMA, Y., TOYA, T., YAMAMURA, T., YAMANAKA, I.,
 YASUNISHI, A., YOSHIDA, K., YOSHINO, M., MURAMATSU, M. and
 HAYASHIZAKI, Y.

Direct Submission
 Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suenhiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)

COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
 further details.

cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in RIKEN
 Genomic Sciences Center and Genome Science Laboratory in RIKEN
 Division of Experimental Animal Research in RIKEN contributed to
 prepare mouse tissues. First strand cDNA was primed with a primer
 [5'-GAGAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTCTTCTT-3'], cDNA was
 prepared by using trihalose thermo-activated reverse transcriptase
 and subsequently enriched for full-length by cap-trapper. cDNA went
 through one round of normalization to Rot = 10.0 and subtraction to
 Rot = 100.0. Second strand cDNA was prepared with the primer
 adapter of sequence [5'-
 GAGAGAGAGATTCGAGTTTATTTATTTATCCCTCCCTCCCTCCCT-3']. cDNA was cleaved
 with BamHI and XhoI. cDNA of size comprised between 3 and 7 kb was
 selected before cloning. Vector: a modified pBluescript KS(+) after
 bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3'
 end: BamHI. Host: DH10B.

FEATURES
 Location/Qualifiers

source

1..3063

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="MGI:1907565"

/db_xref="taxon:10090"

/clone="4932441H21"

/sex="male"

/tissue_type="testis"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="adult"

438..1388

/note="data source:SPR, source key:Q9EPF8, evidence:ISS"

CDS

seq_name: gb_hnc:AK017036

seq_documentation_block:

LOCUS AK017036 1394 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:933433E02:similar to T1 OLFATORY RECEPTOR, full insert sequence.
ACCESSION AK017036
VERSION AK017036.1 GI:12856091
KEYWORDS HTC: CAP trapper.
SOURCE Mus musculus (Strain:C57BL/6J) adult male testis cDNA to mRNA, clone_lib:RIKEN Full-length enriched mouse cDNA library
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

REFERENCE

AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,I., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

REFERENCE

AUTHORS 3 (sites)

TITLE Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujikawa,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsushita,S., Kawaji,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)


```

249  TysrProVal1leYrThrYrIleArgProAlaSerSerYrThrph 265
|||||: : : : : |||||: |||||: : : : :
1085 TAGGGGCTCCACTTCATGATGTCGTGCCACTTCCTCCACACTCC 1134
265  egluArgAspLysValAlaAlaIleuYrThrLeuValThrProThrL 282
|||||: : : : : |||||: : : : : |||||: : : : :
1135 TAGCAAGACAAATATCATCTCTGTCTATACATTTGTCACACACTCC 1184
282  euAsnProMetValTyrSerPheGlnAsnArgLumetGlnAlaGlyLe 298
|||||: : : : : |||||: : : : : |||||: : : : :
1185 TGAACCCCTCATTTACAGCTGAGAAATAGAGAGTGAATGAGCTGT 1234
299  ArgLysVal 301
|||||: : : : :
1235 AGAAGAGTA 1243

seq_name: gb_gss:AQ428256

seq_documentation_block:
LOCUS AQ428256 470 bp DNA linear GSS 24-MAR-1999
DEFINITION CITBI-EI-2578F11.TF CITBI-EI Homo sapiens genomic clone 2578F11,
DNA sequence.
ACCESSION AQ428256
VERSION AQ428256.1 GI:4496022
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 470)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shiyaya,H., Simon,M. and
Venter,J.C.
TITLE Use of BAC End Sequences from Caltech Libraries for Sequence-Ready
Map Building
JOURNAL Unpublished (1997)
COMMENT Other GSSs: CITBI-EI-2578F11.TF
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetlgr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

FEATURES
source 1..470
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CITBI-EI"
/sex="male"
/note="Vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"
BASE COUNT 119 a 122 c 80 g 147 t 2 others
ORIGIN

alignment_scores:
Quality: 531.00 Length: 109
Ratio: 4.963 Gaps: 0
Percent Similarity: 98.165 Percent Identity: 97.248

alignment_block:
US-09-975-308-9 x AQ428256 ..

Align seg 1/1 to: AQ428256 from: 1 to: 470
1 MetAsnHisSerValAlaThrGluPheIleIleuGlyLeuThrLysLys 17

```

```

|||||: : : : : |||||: : : : : |||||: : : : :
142 ATGATATCACAGCCCTTGACTGAGTATATATTGTGGGCTCACCAAAA 191
17  sProGluLeuGlnGlyIleIlePheLeuPhePheLeuIleValTyrLeuV 34
|||||: : : : : |||||: : : : : |||||: : : : :
192 GCGTGAACCTCANGAATATATCTCTCTTTTCTCATTTGTATCTGT 241
34  alaIaPheLeuGlyAsnMetLeuIleIleIleAlaLysIleTyrSerAsn 50
|||||: : : : : |||||: : : : : |||||: : : : :
242 TGGCTTTTTCGGCAACATGCTCATCATTCATTCGCAATATTAACAC 291
51  ThrLeuHisThrProMetYrValPheLeuLeuThrLeuAlaValAs 67
|||||: : : : : |||||: : : : : |||||: : : : :
292 ACCTTGATACGCCCATGATGTTTCTCTCTGACATGGCTGTGTGCA 341
67  pIleIleCysThrThrSerIleIleProLysMetLeuGlyThrMetLeu 84
|||||: : : : : |||||: : : : : |||||: : : : :
342 CATCATCTGCACAGACATCATACCGAAGATGCTGGGACCATGCTAA 391
84  hrSerGluAsnThrThrIleSerTyrAlaIleGlyCysMetSerGlnLeuPheLeu 100
|||||: : : : : |||||: : : : : |||||: : : : :
392 CATCAGAAAATACCATTTATATGACAGCTGCATGCTCCAGCTTCTTG 441
101  PheThrTrpSerLeuGlyAlaGluMet 109
|||||: : : : : |||||: : : : : |||||: : : : :
442 TTCACATGTCCTCTGGGAGCTGAGATG 468

seq_name: gb_gss:A2909618

seq_documentation_block:
LOCUS A2909618 797 bp DNA linear GSS 05-MAR-2001
DEFINITION RPCI-24-222G18.TJ RPCI-24 Mus musculus genomic clone RPCI-24-222G18
, DNA sequence.
ACCESSION A2909618
VERSION A2909618.1 GI:13228563
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 797)
AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akınret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Seq primer: SP6
Class: BAC ends.

FEATURES
source 1..797
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="RPCI-24-222G18"
/sex="male"
/note="Vector: pTRABAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTRABAC1 cloning vector at the

```


14	LeuThlLysLysProCLeuLeuGlnGlyIleIlePheLeuPhePheLeuI	30
792	CTGACCCCAAGTCAAAAGTGAAGCTTGGCTCTTTCCTTTCCT	743
30	eValTYrLeuValAlaPheLeuGlyAsnMetLeuIleIleLeaLal	47
742	GGTGTATGTGACAAACGTGCTGGGAAACCTTCATCATGTGTAC	693
47	LeTYrSerAsnThrIleuHISThr.PrometyrValPheLeuThrIle	63
692	CCGTGTGAGTCCCTCCCTTACACCCCCCAAGTATTTCTTGCTC	643
63	uAlaValAlaSPIleIlecySThrThSerIleIleProLysMetLeu	80
642	GTCGTGTGCTGATATGTGCTTTTCCCTCATCTACGTCCCCCAAG	593
80	LYrHMetLeuThrSerGluAsnThrIleSerTYrAlaGlyCysMetSer	96
592	TAGACCTTCTCTCAGACAAAGAACCATCTTTTCAAGGCTGTCACT	543
97	GlnLeuPheLeuPheThrTrpSerLeuGlyAlaGluMetValLeuPheTh	113
542	CAGATGTTCTCTCTTCACCTTATTTGGGGAGAGGATGTGTTTCTGTG	493
113	rThrMetAlaTYrAspArgTYrValAlaIlecySphProLeuHISLys	130
492	TGTGATGGGCTCTAGATCGATATGTGGCCATCTCTAAGCCCTTGACAT	443
130	erThrIleLeuMetAsnHISLysMetCysValAlaIleLeuLeuSerMetValMet	146
442	TGACCATCATGACCAAGGCGCTCATTTGGGTTTAATTTGTGGCCCTCTGG	393

```
147 AAlaIleAlaValThrAnSerTrpValHisThrAlaLeuIleMetArgLe 163
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
392 GNGGGGGGGCTTTGGCCATCCATAGCAGATTTCTCTTCTGACACT 343
163 unhrphecysglyproasnThrIleasphisphepcysgluileprop 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
342 TCCATTCTGTGGAGCCCATGTTGTGATCTTCTACTGTGATGTCGCC 293
180 rleuLeuAlaLeuSerCysSerProValArgIleasnGluValMetVal 196
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
292 AGGTGATCAAACTGCTGCTACAGACATTTGTTGTTGAGCTGCGTAG 243
197 TyrValAlaAspIleThrLeuAlaIleGlyAspPheIleLeuThrCys 213
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
242 ATTCCAAATATAGCGCTGCTGCTACTGCTGTTGTTGTTGTTGTTG 193
213 eSerTrpGlyPheIleIleValAlaIleLeuArgIleArgThrValGlu 230
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
192 GTCATAC...ACAGTCATCTTATGATGCTATGATCATTCATGAGAGAG 146
230 lYlYsArGlyAlaPheSerThrCysSerSerHisLeuThrValAlaThr 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
145 GCAGGAGAAAGCCATCTCCACCTGACCTCCACATCATCTGTGTCACA 96
247 LeuTyrTrpSerProValIleTyrThrTyrIleArgProAlaSerSer 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
95 CTGCACCTTTGTGCTCCCTGCATCTATGCTATGCGAGCCCTTCACAG 48
263 rThrheGluArgAspLysValAlaAlaLeuTyrThrLeuValThr 279
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
47 ...CTCCCGACGATAAAGCATCTCTGTGTCACCTTCACAGTCATCTCC 3
```

seq_name: gb_uss: A2593814

seq_documentation_block:

LOCUS A2593814 580 bp DNA linear GSS 13-DEC-2000
DEFINITION M0405K03R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0405K03 R, DNA sequence.

ACCESSION A2593814
VERSION A2593814.1 GI:11716004

KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 580)
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausen,A.
and Wright,D., Weis,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb
Plasmid inserts

JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0405 Row: K Column: 03
Seq primer: CACACAGAAACAGCTATGAC
Class: plasmid ends
High quality sequence stop: 580.
Location/Qualifiers
1. 580

FEATURES
source
1. 580
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0405K03"

/clone.lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD22mv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD2 (g1473211419b1AF129072.1), a copy number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 114 a 162 c 129 g 175 t
ORIGIN

alignment_scores:

Quality: 504.00 Length: 179
Ratio: 3.338 Gaps: 0
Percent Similarity: 84.358 Percent Identity: 51.955

alignment_block:

us-09-975-308-9 x A2593814 ..

Align seg 1/1 to: A2593814 from: 1 to: 580

```
129 TySerThrIleMetAnHisMetCysValAlaLeuLeuSerMetVal 145
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1 TACAGCTCTAGATGAGCCCAAGATGTGGGGGCTGAGCCCATGGAGT 50
145 lMetAlaIleAlaValThrAnSerTrpValHisThrAlaLeuIleMet 162
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 ATGTCATCATGTCGTGTGATGATGATGTCACACTGGCTTATATACAC 100
162 rGluThrPheCysGlyProAsnThrIleasphisphepcysgluile 178
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 GCGTGCATCTGTGTGGACCAAGATCATCCACACTTCTCTGTGAGAT 150
179 pProLeuLeuAlaLeuSerCysSerProValArgIleasnGluValMe 195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151 CCCCACCTCTCTGCTCTCTGTGATGCTACATATGTAATACCATAT 200
195 tValTyrValAlaAspIleThrLeuAlaIleGlyAspPheIleLeuThr 212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 GACTCTTTGGAGATTCCTTTTGGAGCGTCATATTTGCTTACT 250
212 ySleSerTrpGlyPheIleIleValAlaIleLeuArgIleArgThrVal 228
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
251 TCGTATCTATGCTGCTGATCATGTCAGCATCTGCGCATGCTTCTGCT 300
229 GluGlyLysArgLysAlaPheSerThrCysSerSerHisLeuThrVal 245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
301 GAGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 350
245 lThrLeuTyrTrpSerProValIleTyrThrTyrIleArgProAlaSer 262
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
351 CTCTGTGCTACTCATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 400
262 eTrpThrPheGluArgAspLysValAlaAlaLeuTyrThrLeuVal 278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
401 GCTACAGTCACAGAAAGCAAGTACTCAGTGTATATACGATCGTC 450
279 ThrProThrLeuAsnProMetValTyrSerPheGlnAsnArgGluMet 295
```


[illegible]

BASE COUNT	276 a	134 c	159 g	171 t
ORIGIN				

alignment_block:
US-09-975-308-9 x AZ607393/rev

Align seg 1/1 to reverse of: AZ607393 from: 1 to: 740

[illegible]

```
seq_documentation_block:
```

LOCUS	AZ553949	676 bp	DNA	linear	GSS-02-OCT-2000
DEFINITION	IM0093J01F Mouse 10kb plasmid UGCG1M library Mus musculus genomic				
CLONE	UUGC1M0093J01 F, DNA sequence.				

VERSION AZ353949.1 GI:10464961

SOURCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus

REFERENCE	AUTHORS
1 (bases 1 to 676)	Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., ...

TITLE Mouse whole genome scaffolding with paired end reads from 10kbb

JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss

University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

Tel: 801 585 5606

Email: ddunn@genetics.utah.edu

Plate: 0093 row: 5 column: 01

Class: plasmid ends
High quality sequence stop: 676.

Source

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U081M0093J01"
 /clone_lib="Mouse 10kb plasmid U081M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-GOLD, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male); was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (<http://www.jax.org/resources/documents/dnares/>). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 173 a 149 c 135 g 219 t
ORIGIN

alignment_scores:
Quality: 493.50 Length: 220
Ratio: 2.920 Gaps: 2
Percent Similarity: 76.818 Percent Identity: 44.545
alignment_block:
US-09-975-308-9 x AZ353949 ..
Align seg 1/1 to: AZ353949 from: 1 to: 676

```
76 ProlysmetleuglythrMetleuthrsergluasnhrilesertryal 92
|||||.....:|||||.....:|||||.....:
18 CCCAAATGATTGTTGACCTTTAGTGAAGAAAAGATATATCTATGT 67
92 aglyCysmetsergluasnhrilesertryal 109
|||||.....:|||||.....:|||||.....:
68 GGGGATGATGTA.CAACCTTTGGGGTCAATTTGGTTCACATGAGA 116
109 etvalleuphethrMetalaTyraPargTyraValaIleCysphe 125
:.....:.....:.....:.....:.....:
117 TCTTCATTCCTTACTGCTGATGCTATGATGATGATGATGATGATG 166
126 ProleuHISTYserThrIleMetasnHISHsmetCysValaIleLeu 142
|||||.....:|||||.....:|||||.....:
167 CCTCCCACTAATGATCATCATGATGATGATGATGATGATGATGAT 216
142 userMetValMetalaIleValaIleAsnserTryValaIleHIST 159
:.....:.....:.....:.....:.....:
217 GCTCGAATGATGATGATGATGATGATGATGATGATGATGATGAT 266
159 euIleMetArgleuthrPheCysGlyProAsnThrIleAspHisPhe 175
|||||.....:|||||.....:|||||.....:
267 TTGTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 316
176 CysgluIleProleuLeuAlaLeuSerCysSerProValArgIleAs 192
|||||.....:|||||.....:|||||.....:
317 TGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 366
192 ngluValMetValIleValaIleAspIlePheLeuAlaIleGlyAsp 209
|||||.....:|||||.....:|||||.....:
367 TGGTGTGTTGTCAGACGACGACGATGATGATGATGATGATGAT 416
209 leleuthrCysIleSerTryGlyPheIleIleValaIleLeuArg 225
|||||.....:|||||.....:|||||.....:
417 TCATCTGCTGATGATGATGATGATGATGATGATGATGATGAT 463
226 ArgThrValIleGluIleValaIleAspIlePheLeuAlaIleGly 242
|||||.....:|||||.....:|||||.....:
464 CAGTCACTCCGACGACGACGACGATGATGATGATGATGATGAT 513
242 urThrValIleThrLeuTyreSerProValIleTryThrTyreArg 259
|||||.....:|||||.....:|||||.....:
514 TGTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 563
259 roIleSerSerTryThrPheGluArgAspIleValaIleAlaLeuTy 275
|||||.....:|||||.....:|||||.....:
564 CTGACACT.....ACCTTCTGACGACGACGATGATGATGATGAT 607
276 ThrleuValIleProThrLeuAsnProMetValIleSerPheGlu 292
|||||.....:|||||.....:|||||.....:
608 ACTATATACCTCCATGCTGATGATGATGATGATGATGATGATG 657
```

292 gclumetin 295
|||||.....
658 AGAAGTAAAG 667

seq_name: gb_gss:BH069789

seq_documentation_block:

LOCUS BH069789 698 bp DNA linear GSS 18-JUL-2001
DEFINITION RPI-24-331A10.TYB RPI-24 Mus musculus genomic clone
RPI-24-331A10, DNA sequence.

ACCESSION BH069789
VERSION BH069789.1 GI:14889386
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 698)
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akintret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPI-24

Unpublished (1999)
Other GSSs: RPI-24-331A10.TYB
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
page: http://www.tigr.org/tldb/Bac_ends/mouse/bac_end_intro.html
Plate: 331 row: A column: 10
Seq primer: T7
Class: BAC ends.

Location/Qualifiers
1..698
/organism="Mus musculus"
/strain="C57BL/6J"
/DB_xref="taxon:10090"
/clone="RPI-24-331A10"
/clone_1b="RPI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: PTARBAC1; site:1: BamHI; site:2: BamHI;
RPI-24 Mouse BAC library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

FEATURES
source

BASE COUNT 230 a 150 c 169 g 149 t
ORIGIN

alignment_scores:
Quality: 483.00 Length: 220
Ratio: 2.825 Gaps: 0
Percent Similarity: 74.348 Percent Identity: 42.174

alignment_block:
US-09-975-308-9 x BH069789/rev ..

Align seg 1/1 to reverse of: BH069789 from: 1 to: 698

```
66 ValAspIleIleCysThrThrSerIleIleProlysmetleuglythr 82
|||||.....:|||||.....:|||||.....:
698 GTGACATCTGCTACCTCTGCTGATCATCTGACGATGATGATGAT 649
82 leuthrsergluasnhrilesertryalaglyCysmetserglu 99
```

```
648 GTGGGACATGGCTCAACACATTTCTCAAGTTCGGCTGCAGTTCATATCT 599
99 heLeuPheThrTrpSerLeuGlyAlaGluMetValLeuPheThrThrMet 115
598 TCCCTTTACCTCTTTCCTTATGCTGCTACCTGCTTGCACATCATG 549
116 AlaTyraProArgTyrValAlaIleCysPheProLeuHisTyrSerThrI 132
548 GCCACACAGCCGCTATGTGGCTGTGCGCACACCATTCCTTATGTACACAT 499
132 eMetAsnHisHisMetCysValAlaLeuLeuSerMetValMetAlaIleA 149
498 CATGACTGAGAAAGCCCGGTAGCTGTGTAACGTGGGACATATGTGGCTG 449
149 IValThrAsnSerTrpValHisThrAlaLeuIleMetArgLeuThrPhe 165
448 GTTTTTCAGTGGCTTATTCGAGCTGTCAAGCCCTTCACCTTTCATTC 399
166 CysGlyProAsnThrIleAspHisPhePheCysGluIleProProluLe 182
398 TGTGGAACAATGATGATCAATTCATATCTGTACCTCCCCCATTTGT 349
182 uAlaLeuSerCysSerProValArgIleAsnGluValMetValTyrValA 199
348 AAAACTGCTATGTGGGACAGCTACATTCAAAGAGTGGTATATGTGTT 299
199 IaAspIleThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyr 215
298 TTGCCATTTTGTTCATGCCCTGCTGTCATGTGTGATCTCTGTGCTCAT 249
216 GlyPheIleIleValAlaIleLeuArgIleArgThrValGluGlyLysAr 232
248 CTGTTTATCTTGTGGCATTTATGCANATCCGCTCAGCTGGAGCGCAGC 199
232 gLysAlaPheSerThrCysSerSerHisLeuThrValValThrLeuTyrT 249
198 CAAAACTTCTACCTGCACCTCCACCTCAGCTGTAAGCTCTCTCT 149
249 ySerProValIleTyrThrTyrIleArgProAlaSerSerTyrThrPhe 265
148 TTGTACTCTTATCTTCATGTAAGAGATATACAGATACAGATCTCTCG 99
266 GluArgAspLysValValAlaAlaLeuTyrThrLeuValThrProThrLe 282
98 GAGAGGACAGAGTTGTGTCTGTCTATACAGGTGTGAGCCCATTAAT 49
282 uAsnProMetValTyrSerPheGlnAsnArgGluMetGln 295
48 GAATCCAGCTCATCTATAGCCTGAGAAATAAAGAGGTAAA 9
```

THIS PAGE BLANK (USPTO)

From: Li, Ruixiang
Sent: Thursday, June 20, 2002 9:56 AM
To: STIC-Biotech/ChemLib
Subject: Sequence search of Application NO: 09/975,308

Please do a standard search on SEQ ID NOS: 8 and 9 against both the commercial and interference nucleic acid databases.

Thank you very much!

Ruixiang Li
GAU 1646
CM1 10E18
Mail Box 10C01
306-0282

Edward Hart
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

BEST AVAILABLE COPY
BEST AVAILABLE COPY
BEST AVAILABLE COPY

RECEIVED
JUN 20 2002
(STIC)

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 6/21/02
Date Completed: 6/26/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: 2
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: X
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)